

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Creación de un Modelo de Recomendación y Guianza
Vocacional Orientado a la Claridad Profesional**

Trabajo de graduación presentado por Oscar Esteban Donis Martínez
para optar al grado académico de Licenciado en Ingeniería en Ciencias
de la Computación y Tecnologías de la Información

Guatemala,

2025

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Creación de un Modelo de Recomendación y Guianza
Vocacional Orientado a la Claridad Profesional**

Trabajo de graduación presentado por Oscar Esteban Donis Martínez
para optar al grado académico de Licenciado en Ingeniería en Ciencias
de la Computación y Tecnologías de la Información

Guatemala,

2025

La creación de este proyecto requirió aplicar muchos de los conocimientos adquiridos a lo largo de mis años de estudio, y de utilizar las habilidades más desarrolladas a lo largo de la carrera, la de aprender e investigar para poder aplicar nuevas tecnologías y enfoques para crear mejores soluciones a problemas ya existentes. Reconozco que el esfuerzo necesario no solo fue el propio, sino de todas las personas que me apoyaron a lo largo de este proceso, soy consciente de que no podría haberlo realizado sin ellos.

Primera y principalmente, quiero agradecer a Dios por haberme proporcionado de todas las herramientas y oportunidades, así como la fuerza, constancia y conocimiento necesario para la realización de este proyecto. Quiero agradecer a mi familia, principalmente, a mis hermanos y padres por su apoyo, comprensión y sacrificios hechos durante el trayecto, no solo de esta tesis, sino de este viaje estudiantil.

Finalmente, quiero agradecer a mi asesor, Ing. Ludwing Cano, por su apoyo e inspiración en la realización de este proyecto, y la paciencia que me brindó durante las diferentes etapas de desarrollo. También quiero agradecer a mi comité examinador, Ing. Alexander Bolaños, por su tiempo y esfuerzo en la revisión de este trabajo, y la paciencia para resolver cualquier duda o inquietud durante los procesos de revisión.

Prefacio	III
Lista de figuras	X
Lista de cuadros	XII
Resumen	XIII
Abstract	XIV
1. Introducción	1
2. Justificación	3
3. Objetivos	5
3.1. Objetivo general	5
3.2. Objetivos específicos	5
4. Alcance	7
4.1. Alcance del proyecto	7
5. Marco teórico	9
5.1. Lenguaje de programación y algoritmo: las herramientas del cómputo	9
5.1.1. Lenguaje de programación	10
5.1.2. Algoritmo	10
5.1.3. Scripts	10
5.2. Inteligencia artificial (IA): la simulación de la cognición humana	10
5.2.1. Definición y alcance	11
5.2.2. Tipos de inteligencia artificial (IA)	11
5.2.3. La revolución de la AI (LLMs)	12
5.2.4. ¿Cómo interpretan el lenguaje los LLMs?	12
5.3. Machine learning (ML): el paradigma del aprendizaje a partir de datos	13
5.3.1. Definición fundamental	13
5.3.2. Modelo de machine learning	14

5.3.3.	El proceso de aprendizaje	14
5.3.4.	Tipos de aprendizaje	14
5.3.5.	Entropía cruzada (función de pérdida)	15
5.3.6.	Hiperparámetros	15
5.3.7.	Validación cruzada	15
5.3.8.	Búsqueda de hiperparámetros	15
5.3.9.	Regularización y sobreajuste	16
5.4.	Análisis y preparación de datos	16
5.4.1.	Limpieza del dataset	16
5.4.2.	Técnicas de imputación	16
5.4.3.	Valores artificiales	16
5.5.	Estadística descriptiva	16
5.5.1.	Variables numéricas	17
5.5.2.	Medidas de tendencia central	17
5.6.	Conceptos clave computacionales	17
5.6.1.	Orquestación centralizada	17
5.6.2.	Arquitectura orientada a servicios	17
5.6.3.	Prompt engineering	17
5.6.4.	Ítem	18
5.6.5.	Instrumento de evaluación	18
5.6.6.	Perfil psicométrico	18
5.7.	Teorías psicológicas	18
5.7.1.	Big five personality	19
5.7.2.	Teoría de holland (Modelo RIASEC)	19
5.7.3.	Teoría del grit	19
5.8.	Factores psicosociales en la elección de carrera del estudiante	19
5.8.1.	La influencia del entorno familiar y las expectativas parentales	20
5.8.2.	El impacto del estatus socioeconómico en el abanico de opciones percibidas	20
5.8.3.	La presión de grupo y las tendencias sociales entre pares	21
5.8.4.	El desarrollo de la identidad vocacional durante la adolescencia	21
5.9.	Características a medir	21
5.9.1.	Habilidades cognitivas	21
5.9.2.	Habilidades no cognitivas	22
5.10.	Evaluaciones de competencias cognitivas y académicas	22
5.10.1.	Programme for the international assessment of adult competencies (PIAAC)	22
5.10.2.	Prueba de aptitud académica (PAA)	22
5.11.	Contexto profesional	23
5.11.1.	Profesión	23
5.11.2.	Carrera profesional	23
5.11.3.	Trayectoria profesional	23
5.11.4.	Orientación vocacional	24
5.12.	Desafíos de la orientación vocacional en la era digital	24
5.12.1.	La paradoja de la información	24
5.12.2.	El futuro del trabajo y habilidades del siglo XXI	25
5.12.3.	La brecha de habilidades (skills gap)	25
5.13.	Modelos y enfoques en la orientación vocacional	26

5.13.1. Modelos tradicionales	26
5.13.2. Enfoques tecnológicos emergentes	27
5.14. Estado del arte de los sistemas de recomendación en contextos educativos (EdRecSys)	27
5.14.1. Recomendación de material	28
5.14.2. Recomendación de carreras profesionales	28
5.15. El rol de los modelos de lenguaje extensos (LLMs) en la guianza personalizada	28
5.15.1. IA conversacional como tutor o consejero	29
5.15.2. Personalización del contenido	29
5.15.3. Retrieval-augmented generation (RAG)	30
5.15.4. Conectar LLMs con bases de datos específicas y acceso a información privada	30
5.16. Componentes específicos del modelo propuesto	32
5.16.1. Motor de recomendación	32
5.16.2. Tipos de motores de recomendación	32
5.16.3. Conjunto de datos (Dataset)	33
5.16.4. División del dataset	33
5.16.5. Feature (característica)	33
5.16.6. Feature engineering	34
5.16.7. Métricas de evaluación	34
5.16.8. Learning-to-Rank (LTR)	35
5.17. Herramientas utilizadas	35
5.17.1. Python	35
5.17.2. Librería pandas	35
5.17.3. Librería scikit-learn	36
5.17.4. Psytoolkit	36
5.17.5. XGBoost	36
5.17.6. Vector DB	36
5.17.7. Bases de datos no SQL	37
5.18. Consideraciones éticas y desafíos de la IA en la orientación vocacional	38
5.18.1. Sesgos algorítmicos	38
5.18.2. Privacidad de los datos del estudiante	39
5.18.3. El problema de la caja negra (black box)	39
6. Metodología	40
6.1. Diseño del sistema	40
6.1.1. Justificación del desacoplamiento	40
6.1.2. Arquitectura del sistema	41
6.1.3. Diagrama de la arquitectura del sistema	42
6.2. Módulo de información	42
6.2.1. Recopilación de datos	43
6.2.2. Limpieza de datos	43
6.2.3. Atributos y entidades de datos	44
6.2.4. Comparativas de herramientas disponibles	45
6.3. Módulo de guianza vocacional	47
6.3.1. Necesidades del sistema	47
6.3.2. Requisitos de seguridad y ética	47
6.3.3. Comparativa de herramientas disponibles	48

6.4.	Módulo de recomendación	50
6.4.1.	Definición y recolección de datos	50
6.4.2.	Modelo de recomendación de carreras (learning-to-rank)	51
6.4.3.	Algoritmo de recomendación de contenido (filtro basado en contenido)	51
6.4.4.	Comparativa de herramientas disponibles	52
6.5.	Módulo de comunicación	55
6.5.1.	Arquitectura y principios de diseño	55
6.5.2.	Necesidades funcionales y flujos de datos	55
6.5.3.	Comparativa de herramientas disponibles	56
7.	Resultados	59
7.1.	Resultados encuesta inicial	59
7.1.1.	Perfil de usuarios encuestados	59
7.1.2.	Exploración y apoyo en futuro profesional	63
7.1.3.	Búsqueda de información y recursos	65
7.1.4.	Expectativas sobre herramientas de orientación vocacional	66
7.2.	Módulo de información (obj. específicos 1 y 2 - 3.2)	70
7.2.1.	Justificación de elecciones técnicas	70
7.2.2.	Desarrollo	72
7.2.3.	Atributos y entidades de datos documentales	73
7.2.4.	Estructuras y modelos de datos vectoriales	74
7.2.5.	Pruebas funcionales	76
7.2.6.	Resultados de la prueba	77
7.3.	Módulo de guianza vocacional (obj. específicos 3 y 4 - 3.2)	78
7.3.1.	Justificación de elecciones técnicas	79
7.3.2.	Desarrollo	80
7.3.3.	Pruebas funcionales	82
7.3.4.	Resultados de la prueba	83
7.4.	Módulo de recomendación (obj. específicos 5, 6 y 7 - 3.2)	84
7.4.1.	Justificación de elecciones técnicas	84
7.4.2.	Desarrollo	85
7.4.3.	Pruebas funcionales	98
7.4.4.	Resultados de la prueba	100
7.5.	Módulo de comunicación (obj. específicos 8, 9 y 10 - 3.2)	101
7.5.1.	Justificación de elecciones técnicas	101
7.5.2.	Desarrollo	103
7.5.3.	Arquitectura de la API	103
7.5.4.	Endpoints de la API	103
7.5.5.	Flujo de comunicación entre modelos y bases de datos	104
7.5.6.	Pruebas funcionales	105
7.5.7.	Resultados de la prueba	106
8.	Conclusiones	108
8.1.	Encuesta inicial	108
8.2.	Módulo de información (obj. específicos 1 y 2 - 3.2)	109
8.3.	Módulo de guianza vocacional (obj. específicos 3 y 4 - 3.2)	109
8.4.	Módulo de recomendación (obj. específicos 5, 6 y 7 - 3.2)	110
8.5.	Módulo de comunicación (obj. específicos 8, 9 y 10 - 3.2)	111

8.6. Conclusión final	112
9. Recomendaciones	113
9.1. Recomendaciones para el sistema de recomendación	113
9.1.1. Encuesta inicial	113
9.1.2. Módulo de información	113
9.1.3. Módulo de guianza vocacional	113
9.1.4. Módulo de recomendación	114
9.1.5. Módulo de comunicación	114
10. Bibliografía	116
11. Anexos	129
11.1. Resultados	129
11.1.1. Resultados módulo de información	129
11.1.2. Resultados módulo de comunicación	146
11.1.3. Resultados módulo de guianza vocacional	150
11.1.4. Resultados módulo de recomendación	171
11.1.5. Resultados módulo de comunicación	179

Lista de figuras

1.	Diagrama del ciclo de aprendizaje iterativo de un modelo de Machine Learning	14
2.	Diagrama del proceso de aprendizaje de un modelo de Machine Learning . . .	20
3.	Diagrama de Retrieval-Augmented Generation (RAG)	30
4.	Diagrama de la arquitectura del sistema propuesta	42
5.	Distribución de género de los usuarios encuestados	60
6.	Distribución de edades de los usuarios encuestados	60
7.	Distribución de etnicidad de los usuarios encuestados	61
8.	Distribución de grados de los usuarios encuestados	61
9.	Distribución de localización de centros de estudios de los usuarios encuestados	62
10.	Distribución de marca de teléfono de los usuarios encuestados	62
11.	Distribución primera pregunta de apoyo futuro de los usuarios encuestados . .	63
12.	Distribución segunda pregunta de apoyo futuro de los usuarios encuestados .	64
13.	Distribución tercera pregunta de apoyo futuro de los usuarios encuestados . .	64
14.	Distribución primera pregunta de búsqueda de recursos de los usuarios encuestados	65
15.	Distribución segunda pregunta de búsqueda de recursos de los usuarios encuestados	66
16.	Distribución tercera pregunta de búsqueda de recursos de los usuarios encuestados	66
17.	Distribución primera pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados	67
18.	Distribución segunda pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados	67
19.	Distribución tercera pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados	68
20.	Distribución cuarta pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados	69
21.	Distribución quinta pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados	69
22.	Diagrama de la arquitectura del sistema de guianza vocacional	80
23.	Participantes del test hecho en la plataforma PsyToolkit	89

24.	Cantidad de participantes recolectados por carrera	90
25.	Flujo de ejecución del modelo de recomendación	94
26.	Flujo del algoritmo de recomendación de contenido	97
27.	Diagrama de la arquitectura de la API de comunicación	103
28.	Diagrama de flujo RAG - Pregunta del usuario hasta respuesta del LLM . . .	104
29.	Resultados de la ejecución de la prueba del objetivo de la base de datos . . .	135
30.	PIAAC - Unidad 817: Pregunta 1	141
31.	PIAAC - Unidad 826: Pregunta 2	142
32.	PIAAC - Unidad 826: Pregunta 3	142
33.	PAA - Entendimiento del inglés: Pregunta 22	145
34.	PAA - Razonamiento matemático: Pregunta 17	146
35.	Resultados de la ejecución de la prueba del objetivo de la guianza	171
36.	Scripts creados para las pruebas	171
37.	Resultados de la ejecución de la prueba del objetivo de la recomendación . . .	172
38.	Resultados de la ejecución de la prueba del objetivo de la recomendación . . .	172
39.	Resultados de la ejecución de la prueba del objetivo de la recomendación . . .	173
40.	Resultados de la ejecución de la prueba del objetivo de la recomendación . . .	174
41.	Resultados de la ejecución de la prueba del objetivo de la recomendación . . .	175
42.	Resultados de la ejecución de la prueba del objetivo de la recomendación . . .	176
43.	Resultados de la ejecución de la prueba del objetivo de la recomendación . . .	177
44.	Tareas realizadas para la prueba del módulo de comunicación	179
45.	Aspectos evaluados para la prueba del módulo de comunicación	180
46.	Métricas de la prueba 1 del módulo de comunicación	180
47.	Métricas de la prueba 2 del módulo de comunicación	181
48.	Métricas de la prueba 3 del módulo de comunicación	182
49.	Métricas de la prueba 4 del módulo de comunicación	183
50.	Métricas de la prueba 5 del módulo de comunicación	184
51.	Métricas de la prueba 6 del módulo de comunicación	185

Lista de cuadros

1.	Comparativa: SQL vs NoSQL	45
2.	Comparativa: Bases de datos no relacionales	46
3.	Comparativa: Bases de datos vectoriales	46
4.	Comparativa: RAG vs. fine-tuning	48
5.	Comparativa: modelos de embeddings	49
6.	Comparativa: modelos de LLMs	50
7.	Comparativa de lenguajes de programación para machine learning	52
8.	Comparativa de varios algoritmos de <i>machine learning</i>	53
9.	Comparativa de algoritmos de árboles de decisión	54
10.	Comparación de las arquitecturas REST y GraphQL	56
11.	Comparativa de lenguajes de programación para la implementación de la API	57
12.	Comparativa de frameworks TypeScript para APIs RESTful	58
13.	Resultados de la prueba de integridad estructural	78
14.	Resultados de la prueba de precisión de consulta	78
15.	Resultados generales de la prueba	83
16.	Criterios de aceptación para el módulo de recomendación	99
17.	Resultados de la ejecución de la prueba del objetivo de la recomendación	100
18.	Resultados de la prueba del objetivo de la recomendación	100
19.	Resultados promediados de la prueba funcional por módulo: códigos E2/E1/E0 y satisfacción (escala 1–5)	106
20.	Resultados de la pregunta 1	151
21.	Resultados de la pregunta 2	152
22.	Resultados de la pregunta 3	153
23.	Resultados de la pregunta 4	154
24.	Resultados de la pregunta 5	155
25.	Resultados de la pregunta 6	156
26.	Resultados de la pregunta 7	157
27.	Resultados de la pregunta 8	158
28.	Resultados de la pregunta 9	159
29.	Resultados de la pregunta 10	160
30.	Resultados de la pregunta 11	161

31.	Resultados de la pregunta 12	162
32.	Resultados de la pregunta 13	164
33.	Resultados de la pregunta 14	165
34.	Resultados de la pregunta 15	166
35.	Resultados de la pregunta 16	167
36.	Resultados de la pregunta 17	168
37.	Resultados de la pregunta 18	169
38.	Resultados de la pregunta 19	169
39.	Resultados de la pregunta 20	171
40.	Ejemplo de datos de prueba para el objetivo de la recomendación	178

En la actualidad, los estudiantes de bachillerato enfrentan dificultades para elegir una profesión. Esta decisión es cada vez más importante con problemas como la movilidad y la deserción universitaria cada vez más presentes. La investigación autónoma dentro de la vasta cantidad de información del internet puede ser abrumadora, gracias al ordenamiento desestructurado dificulta la búsqueda de información relevante. Las inteligencias artificiales (IA) generativas, por su parte, no ofrecen un valor significativo, al carecer del contexto personalizado del usuario y de una base de conocimiento limitado a la información de su entrenamiento o a una búsqueda web superficial. La solución propuesta comprende un modelo de recomendación y guianza vocacional que utilizan técnicas de inteligencia artificial y de recuperación de información dinámica para proporcionar respuestas y recomendaciones personalizadas sobre carreras universitarias ofrecidas en la Universidad del Valle de Guatemala, utilizando el perfil del usuario.

En la creación del perfil del usuario, se implementó una prueba de evaluación que permite obtener información relevante del usuario, como su personalidad, consistencia e intereses vocacionales. Para proveer información actualizada y confiable, se implementó una base de datos que contiene información relevante de 33 carreras universitarias ofrecidas por la Universidad del Valle de Guatemala, con un 100 % de precisión en la recuperación de información actualizada hasta la fecha de la realización del proyecto. Además, se implementó un sistema de consulta basado en un Large Language Model (LLM), con una satisfacción comprobada del 97.78 %, que utiliza tanto información de la base de datos como el perfil del usuario para proporcionar respuestas. Complementando la solución, se implementó un prototipo de un sistema de recomendación, con una viabilidad demostrable del 100 % en su ejecución, que permite la recomendación de carreras universitarias y contenido académico compatibles con el perfil del usuario.

Estos resultados nos permiten explorar la efectividad y el impacto, que la combinación del contexto del usuario en diferentes áreas con la utilización de nuevas tecnologías utilizadas en el mercado actual como los modelos de LLM, las base de datos vectoriales y la generación aumentada por recuperación, produce en la orientación vocacional de estudiantes de bachillerato.

Currently, high school students face difficulties in choosing a profession. This decision is increasingly important given the growing problems of student mobility and university dropout rates. Independent research within the vast amount of information on the internet can be overwhelming, as its unstructured organization makes it difficult to find relevant information. Generative artificial intelligence (AI), by its nature, does not offer significant value, lacking the user's personalized context and having a knowledge base limited to its training information or a superficial web search. The proposed solution involves a vocational guidance and recommendation model that uses artificial intelligence and dynamic information retrieval techniques to provide personalized answers and recommendations about university programs offered at Universidad del Valle de Guatemala, based on the user's profile.

To create the user profile, an assessment test was implemented to obtain relevant information about the user, such as their personality, consistency, and vocational interests. To provide up-to-date and reliable information, a database was implemented containing relevant information on 33 university programs offered by the Universidad del Valle de Guatemala, with 100 % accuracy in retrieving information updated to the date of the project. In addition, a query system based on a Large Language Model (LLM) was implemented, with a proven satisfaction rate of 97.78 %. This system uses both database information and the user's profile to provide answers. Complementing the solution, a prototype recommendation system was implemented, with demonstrable 100 % feasibility in its execution. This system allows recommendations of university programs and academic content compatible with the user's profile.

These results allow us to explore the effectiveness and impact of combining the user's context in different areas with the use of new technologies currently implemented in the market, such as: LLM models, vector databases, and augmented generation by retrieval, in the vocational guidance of high school students.

La elección de una carrera profesional es una decisión fundamental que marca la vida adulta. Si bien algunas personas ya lo tienen definido desde una edad temprana, la gran mayoría de los estudiantes enfrentan esta etapa con múltiples opciones, más ninguna idea concreta. Esta tarea se vuelve más compleja dado que no se puede elegir ninguna profesión sin la respectiva exposición a temas y conceptos que cada una engloba, como nos dice [1]: un 36.8 % de los estudiantes encuestados afirmó que la escuela no les proporcionó alguna clase de herramienta vocacional. Por ello un número considerable de estudiantes tienden a cambiar o abandonar sus estudios superiores debido a expectativas irreales que llegan antes de ingresar a la carrera, esto nos lo expresa mejor [2]: al exponernos que la segunda de las razones por las que los estudiantes abandonan sus estudios superiores es por pérdida de interés en la carrera. Esta realidad no siempre resulta en el cambio o abandono de la carrera; algunas personas logran finalizar sus estudios, pero descubren posteriormente que no disfrutaban laborar dentro del área que eligieron, como nos dice [3] : una de las causas de la desmotivación de la baja activación de nuestras metas, que no pueden ser llenadas dentro de la situación laboral actual. Lo que genera una baja disposición a realizar tareas, bajando así nuestro rendimiento. Esto puede causar dificultad para conseguir empleo o sufrir de depresión y mal rendimiento, asociados a la falta de motivación y realización en sus labores.

Idealmente, la solución implicaría encontrar exponer a los alumnos a una amplia variedad de campos y materias. Sin embargo, esto representa todo un desafío, ya que brindar conocimiento y experiencia de cada campo a cada estudiante se requiere de una gran variedad de profesionales y el tiempo de interacción individual sería insuficiente para causar un impacto significativo. Por otro lado, si bien el internet proporciona toda clase de conocimiento, presenta sus propios desafíos, asemejándose a una biblioteca sin guía sobre qué estudiar o aprender para dominar alguna disciplina. Gracias al internet, los recursos están disponibles, pero el verdadero problema radica en la orientación necesaria para mantener a los estudiantes motivados, y mitigar el estrés y la incertidumbre del proceso. A esto se suma la limitación inherente de los orientadores vocacionales en cuestión de temas y conocimiento general sobre la mayoría de campos.

Frente a este panorama, la solución ideal parecería ser una entidad que tenga la capacidad de guiar a los estudiantes en cualquier tema o campo, ofreciendo recomendaciones adaptadas a la persona utilizando su personalidad y habilidades. Dadas las características especificadas, e irreales hasta cierto punto, se planea explorar una solución utilizando inteligencia artificial (IA). Esta tecnología, implementada adecuadamente, tiene el potencial de proporcionar una experiencia personalizada a cada alumno, superando las limitaciones de conocimiento, tiempo y disponibilidad mencionadas.

El objetivo del presente trabajo es explorar el diseño y la implementación de un modelo de inteligencia artificial que permita brindar orientación vocacional personalizada a los estudiantes, sin tener las limitaciones previamente expuestas que las soluciones actuales presentan para las nuevas generaciones en el contexto de un mundo laboral en constante cambio. Para ello, se propone la creación de un dataset de entrenamiento que contendrá información anonimizada de diferentes usuarios (resultados de pruebas de personalidad y habilidades, carreras elegidas y su nivel de satisfacción). Este dataset será utilizado para el entrenar un modelo capaz de proporcionar recomendaciones de carreras y temas relevantes, guiando al estudiante a probar a explorar nuevas opciones y asegurando una elección más informada sobre su futuro profesional.

La presente investigación se justifica ante la significativa incertidumbre que enfrentan los estudiantes de bachillerato al momento de elegir una profesión. Esta decisión constituye uno de los hitos más trascendentales en su desarrollo personal y profesional, y cuya dificultad se evidencia en las altas tasas de movilidad y deserción universitaria. Estudios recientes indican que entre el 30 % y el 40 % de los universitarios cambian de carrera durante los primeros semestres [4], y que aproximadamente entre el 7.5 % y 8.5 % abandona sus estudios superiores en el primer año [5]. Este trabajo propone el desarrollo de una herramienta de orientación vocacional para mitigar esta problemática, especialmente en un contexto global volátil que exige decisiones profesionales cada vez más informadas.

Uno de los factores principales que origina esta indecisión es la limitada exposición de los estudiantes a diversas áreas del conocimiento y habilidades prácticas. La solución ideal implicaría guiar a cada alumno a través de experiencias interactivas en los distintos campos profesionales; sin embargo, esta alternativa enfrenta un desafío logístico insuperable por la falta de suficientes profesionales y mentores para atender la demanda estudiantil.

Si bien herramientas como internet ofrecen acceso a un vasto cúmulo de información, esta se encuentra a menudo desestructurada y carece del acompañamiento necesario para una profundización efectiva en los temas. La investigación autónoma, además, puede resultar un proceso solitario y desmotivador para el estudiante, dificultando la constancia en el aprendizaje. Las inteligencias artificiales (IA) genéricas, por su parte, tienden a ofrecer un valor similar al de una búsqueda web estándar, al carecer de un contexto personalizado del usuario y de una base de conocimiento especializada en orientación vocacional.

No obstante, la tecnología de IA posee el potencial para convertirse en la guía personalizada idónea. Una IA debidamente entrenada puede ofrecer disponibilidad 24/7, un profundo conocimiento en múltiples áreas y una capacidad incansable para explicar conceptos y resolver dudas. Por lo tanto, se plantea el diseño de un modelo de IA con la capacidad de orientar a los estudiantes en la exploración de carreras y temas afines, permitiéndoles descubrir vocaciones acordes a sus habilidades y preferencias de manera autónoma pero guiada.

Este enfoque permitiría: 1) asegurar una exploración vocacional más personalizada y estructurada; 2) motivar a los estudiantes a considerar trayectorias profesionales que previamente no contemplaban; 3) facilitar la toma de decisiones informadas; y 4) preparar a los jóvenes para interactuar con profesionales del sector con un conocimiento previo y preguntas específicas, reduciendo así el estrés asociado a esta etapa.

Para que este modelo de IA sea accesible y efectivo, debe integrarse en una plataforma digital, como una aplicación web o un entorno educativo virtual. La elección de un formato de aprendizaje en línea (e-learning) no es trivial, ya que puede “aumentar la retención de estudiantes y empleados hasta en un 50 %” [6]. Esta sinergia entre un modelo de IA especializado y una plataforma de e-learning optimizada puede potenciar la adopción de la herramienta y la consistencia en el proceso de aprendizaje, consolidando una solución robusta y eficaz para la orientación vocacional.

3.1. Objetivo general

Diseñar e implementar un sistema de recomendación y guianza vocacional basado en inteligencia artificial, con el propósito de orientar a los estudiantes en la exploración de carreras universitarias mediante información confiable, actualizada y contextualizada.

3.2. Objetivos específicos

1. Diseñar y construir una base de conocimiento que catalogue y estructure la información de al menos 33 carreras universitarias ofrecidas por la Universidad del Valle de Guatemala, incluyendo un mínimo de cinco atributos esenciales por carrera.
2. En la fase de validación se espera una precisión semántica del 95 %, en la ejecución de al menos 30 pruebas de consulta, asegurando que las consultas retornen información correcta, coherente y relevante con respecto al conocimiento almacenado.
3. Implementar un sistema de guianza vocacional basado en un Modelo de Lenguaje Extenso (LLM) que utilice la base de conocimiento construida para responder preguntas sobre carreras universitarias.
4. Durante la fase de validación, se espera una fidelidad de información y relevancia del 85 %, en la ejecución de un conjunto de al menos 20 consultas.
5. Generar perfiles de usuario multidimensionales mediante la aplicación de una batería de pruebas psicométricas seleccionadas para medir rasgos determinantes de un perfil.
6. Desarrollar e implementar el prototipo de un sistema de recomendación capaz de generar sugerencias de carreras universitarias y contenido académico basado en el perfil del usuario.

7. Durante la fase de validación, se ejecutarán pruebas funcionales del flujo de recomendación inicial utilizando al menos 20 perfiles de usuario simulados, garantizando una tasa de éxito de ejecución igual o superior al 95 %.
8. Implementar un sistema orquestador que maneje la comunicación entre los módulos del sistema.
9. Durante la fase de validación, se implementará al menos 10 pruebas que midan la funcionalidad y la satisfacción del sistema general, con una muestra de al menos 5 usuarios.
10. Garantizar un éxito en la ejecución de las tareas superior al 90 % y una satisfacción por parte de los usuarios mayor al 80 % en las pruebas del sistema General.

4.1. Alcance del proyecto

El presente trabajo de graduación se enfoca en el diseño, implementación y validación de un prototipo funcional para un sistema de recomendación y guianza vocacional basado en inteligencia artificial. El sistema se estructura mediante una arquitectura modular desacoplada que favorece la independencia funcional de sus componentes, organizados en cuatro módulos principales: *información*, *guianza vocacional*, *recomendación* y *comunicación*.

Alcances del proyecto (inclusiones)

Módulo de información. se desarrollará una base de conocimiento centralizada que catalogará la información académica y profesional de treinta y tres (33) carreras universitarias ofrecidas por la Universidad del Valle de Guatemala (UVG). Los datos se almacenarán en una base de datos NoSQL (*DocumentDB*) y en una base de datos vectorial (*Pinecone*), permitiendo la recuperación semántica de información.

Módulo de guianza (chatbot). se implementará un sistema conversacional que empleará un Modelo de Lenguaje Extenso (LLM) bajo una arquitectura *Retrieval-Augmented Generation* (RAG) para responder preguntas de los usuarios. Las respuestas se generarán únicamente a partir del contenido de la base de conocimiento de la UVG. Se utilizará el modelo *Gemini 2.0 Flash* para la generación y *gemini-embedding-001* para la vectorización de documentos.

Módulo de recomendación (prototipo). se diseñará y entrenará un prototipo de modelo de aprendizaje automático basado en el algoritmo *Gradient Boosting (XGBoost)*, con el objetivo de demostrar la viabilidad técnica de recomendar carreras universitarias y contenido académico. El perfil del usuario se construirá a partir de instrumentos psicométricos que evalúan la personalidad (*Big Five*), los intereses (*RIASEC*), la perseverancia (*Grit*) y las habilidades cognitivas (basadas en *PIAAC* y *PAA*).

Módulo de comunicación (backend). el entregable principal consistirá en un backend funcional expuesto a través de una API *RESTful*. Dicha API, desarrollada en *TypeScript* mediante el framework *Hono*, orquestará la comunicación entre los diferentes módulos y permitirá la integración futura con aplicaciones web o móviles.

Limitaciones del proyecto (exclusiones)

Datos de entrenamiento del modelo. la principal limitación del proyecto radica en la recolección de datos para el módulo de recomendación. No se alcanzó la muestra mínima esperada de seiscientos sesenta (660) participantes reales, logrando únicamente sesenta y un (61) resultados utilizables. En consecuencia, el modelo se entrenará con datos sintéticos (*dummy data*) con el fin de validar la arquitectura. El proyecto se concibe como una *Prueba de Concepto* (*Proof of Concept*, PoC) orientada a demostrar la factibilidad técnica del flujo de recomendación, sin generar predicciones validadas para usuarios reales.

Alcance de la información. la base de conocimiento se limitará exclusivamente a las treinta y tres (33) carreras de la Universidad del Valle de Guatemala. No se incluirán programas académicos de otras universidades ni se realizará un análisis exhaustivo del mercado laboral fuera de las fuentes documentales consultadas.

Factores psicosociales. el modelo no considerará variables psicosociales complejas, tales como la influencia familiar, el estatus socioeconómico o la presión de grupo, debido a la dificultad de su medición sistemática en el contexto del presente estudio.

Interfaz de usuario (frontend). no se desarrollará una interfaz gráfica para el usuario final. Las pruebas y validaciones se ejecutarán mediante la API *RESTful* del sistema.

Despliegue y mantenimiento. el alcance del proyecto se limita al diseño, implementación y pruebas funcionales del prototipo. No se contempla el despliegue a un entorno de producción ni la definición de planes de mantenimiento o escalabilidad a largo plazo.

Síntesis del alcance

En conjunto, este proyecto establece las bases conceptuales y técnicas para un sistema de orientación vocacional inteligente, demostrando la viabilidad de integrar modelos de lenguaje y algoritmos de recomendación en un entorno educativo. Si bien su alcance se limita a una prueba de concepto, constituye un punto de partida sólido para investigaciones futuras que busquen integrar datos reales y ampliar la cobertura institucional. El desarrollo de la interfaz interactiva para el usuario final y la implementación de la API en un entorno de producción, son características tratadas en otros trabajos de graduación.

La elección de una carrera profesional es una de las decisiones más significativas en la vida de una persona, con implicaciones profundas en su desarrollo personal y socioeconómico. Tradicionalmente, este proceso ha sido guiado por orientadores vocacionales, cuyo consejo se basa en una combinación de experiencia, conocimiento del mercado laboral y la aplicación de pruebas psicométricas. Sin embargo, en un mundo cada vez más complejo y digitalizado, caracterizado por una sobreabundancia de información y la rápida evolución de las profesiones, los enfoques tradicionales enfrentan nuevos desafíos.

Este capítulo presenta el marco teórico que sustenta el desarrollo de un sistema de recomendación de carreras basado en inteligencia artificial. El objetivo es doble: por un lado, fundamentar la necesidad de un enfoque data-driven para la orientación vocacional y, por otro, establecer las bases conceptuales, psicológicas y tecnológicas sobre las que se construye el modelo propuesto. Se argumenta que la sinergia entre la psicometría clásica y las técnicas avanzadas de aprendizaje automático puede ofrecer una solución más personalizada, objetiva y escalable a los desafíos de la orientación vocacional contemporánea.

5.1. Lenguaje de programación y algoritmo: las herramientas del cómputo

Para comprender el funcionamiento de un sistema de recomendación de carreras basado en inteligencia artificial, es primordial establecer una base sólida sobre los conceptos que sustentan su desarrollo. Esta sección introduce los lenguajes de programación y los algoritmos, los pilares fundamentales sobre los que se construye cualquier software, incluyendo las complejas arquitecturas de los sistemas inteligentes.

5.1.1. Lenguaje de programación

Un **lenguaje de programación** es el conjunto de instrucciones que permiten el traslado de comandos a una máquina [7]. Su función primordial es servir como un puente entre el pensamiento humano, que es abstracto y conceptual, y el lenguaje binario de las máquinas, que es concreto y operativo. Los programadores utilizan estos lenguajes para escribir código fuente, una forma legible para el ser humano que, mediante un proceso de compilación o interpretación, se traduce en instrucciones ejecutables por el procesador del ordenador [8].

5.1.2. Algoritmo

Si el lenguaje de programación es el vocabulario y la gramática, el algoritmo es la narrativa o la receta. Se define como: «conjunto ordenado y finito de instrucciones o pasos definidos que permiten resolver un problema o realizar una tarea específica» [9]. También se puede describir como una entidad abstracta que describe un procedimiento paso a paso, independientemente del lenguaje en que se implemente. Debe ser preciso, no ambiguo y garantizar la terminación tras un número finito de pasos [10].

El término deriva del nombre del matemático persa del siglo IX, Muhámmad ibn Mūsā al-Khwārizmī, cuyo trabajo en álgebra y sistemas numéricos sentó las bases para los procedimientos sistemáticos de resolución de problemas. En la ciencia de la computación moderna, los algoritmos son el núcleo de cualquier programa. Desde una simple ordenación de una lista hasta el complejo entrenamiento de una red neuronal, todo se rige por un algoritmo subyacente [11].

5.1.3. Scripts

En computación, un *script* es una secuencia de instrucciones —generalmente de un lenguaje interpretado— que automatiza una tarea o proceso que de otro modo se ejecutaría manualmente [12].

5.2. Inteligencia artificial (IA): la simulación de la cognición humana

Una vez sentadas las bases computacionales, es crucial adentrarse en el campo de la Inteligencia Artificial (IA), el motor conceptual que impulsa el sistema de recomendación propuesto. Esta sección define qué es la IA, su alcance y sus diferentes tipos, proporcionando el contexto necesario para entender cómo una máquina puede emular procesos cognitivos humanos para analizar perfiles de estudiantes y sugerir trayectorias profesionales personalizadas.

La **Inteligencia Artificial (IA)** es un campo expansivo de la informática que se ocupa de la creación de máquinas y sistemas capaces de realizar tareas que, tradicionalmente, han requerido de la inteligencia humana. Estas tareas incluyen el razonamiento, el aprendizaje,

la resolución de problemas, la percepción, la comprensión del lenguaje y la toma de decisiones [13].

5.2.1. Definición y alcance

La **Inteligencia Artificial** se puede definir como «el estudio y diseño de agentes inteligentes, donde un agente es cualquier entidad que percibe su entorno a través de sensores y actúa sobre él mediante actuadores» [14]. Bajo esta perspectiva, el objetivo de la IA es construir agentes que actúen de forma racional, es decir, que seleccionen la acción que se espera que maximice una medida de rendimiento, dadas las evidencias proporcionadas por la secuencia de percepciones y cualquier conocimiento previo que el agente posea. La IA abarca «lógica, probabilidad y matemáticas continuas; percepción, razonamiento, aprendizaje y acción; equidad, confianza, bien social y seguridad» [14].

5.2.2. Tipos de inteligencia artificial (IA)

La clasificación más común de la IA se basa en la capacidad y funcionalidad de los sistemas [13].

- **Máquinas reactivas:** ia limitada que solo reacciona a diferentes tipos de estímulos según reglas preprogramadas. No utiliza memoria y, por lo tanto, no puede aprender con nuevos datos. Deep Blue de IBM, que venció al campeón de ajedrez Garry Kasparov en 1997, fue un ejemplo de máquina reactiva.
- **Memoria limitada:** la mayoría de las IA modernas se consideran de memoria limitada. Pueden usar la memoria para mejorar con el tiempo al ser entrenadas con nuevos datos, normalmente a través de una red neuronal artificial u otro modelo de entrenamiento. El aprendizaje profundo, un subconjunto del aprendizaje automático, se considera inteligencia artificial de memoria limitada.
- **Teoría de la mente:** la IA con teoría de la mente no existe actualmente, pero se está investigando sobre sus posibilidades. Describe una IA que puede emular la mente humana y tiene capacidades de toma de decisiones iguales a las de un humano, incluyendo reconocer y recordar emociones y reaccionar en situaciones sociales como lo haría una persona.
- **Autoconciencia:** un paso más allá de la IA con teoría de la mente, la IA autoconsciente describe una máquina mítica que es consciente de su propia existencia y tiene las capacidades intelectuales y emocionales de un ser humano. Al igual que la IA con teoría de la mente, la IA autoconsciente no existe actualmente.

Los tipos de IA que se estarán utilizando para este proyecto serán los de máquinas reactivas y memoria limitada.

5.2.3. La revolución de la AI (LLMs)

Un Large Language Model (LLM) es un sistema que aprende a predecir la siguiente palabra o “pieza” de texto a partir del contexto anterior. Para hacerlo, se entrena con enormes colecciones de texto y se apoya en la arquitectura Transformer, cuyo mecanismo central - la autoatención- permite que el modelo “mire” simultáneamente diferentes partes del texto para capturar relaciones de largo alcance y generar respuestas coherentes. Esta arquitectura sustituyó las redes recurrentes/capas convolucionales en muchas tareas de lenguaje por su eficiencia y calidad [15].

5.2.4. ¿Cómo interpretan el lenguaje los LLMs?

El lenguaje de los LLM (tokens)

En modelado de lenguaje, un token es la unidad mínima de texto con la que opera un modelo (p. ej., un LLM) durante el entrenamiento y la inferencia. Dependiendo del método de tokenización, un token puede ser un carácter, una palabra o —más comúnmente hoy— una sub-palabra (segmentos como “inter-”, “-ción”), lo que reduce el vocabulario y permite manejar palabras raras o no vistas. Técnicas populares incluyen WordPiece, BPE (Byte-Pair Encoding), byte-level BPE y Unigram LM / SentencePiece, que segmentan texto para equilibrar cobertura y compacidad del vocabulario. En arquitecturas Transformer, los tokens se embeben en vectores (embeddings) con positional encodings y se procesan con auto-atención para predecir la probabilidad del siguiente token; por ello, el largo del contexto y el costo de uso de un LLM suelen medirse en tokens. En suma, “token” es el átomo operativo del texto en LLMs: define cómo se representa el texto, cuánto contexto cabe, y cómo se computa la probabilidad y el costo por uso.

Las letras de los LLM (embeddings)

Un embedding es una representación numérica de baja dimensión que mapea objetos discretos (p. ej., palabras, frases, nodos o ítems) a vectores en un espacio continuo de forma que proximidad geométrica = similitud semántica/funcional. En NLP, los embeddings “estáticos” como word2vec y GloVe capturan regularidades léxicas a partir de co-ocurrencias, mientras que los embeddings contextuales (p. ej., BERT) dependen del contexto y mejoran el desempeño en tareas aguas abajo; para oraciones, variantes como Sentence-BERT optimizan directamente la similitud semántica de sentencias. En términos generales, los embeddings son parte de la aprendizaje de representaciones, que busca codificaciones útiles para resolver tareas posteriores (búsqueda semántica, recuperación RAG, clustering, recomendación, etc.) [16] [17] [18] [19] [20].

Cómo funcionan (alto nivel)

- **Tokenización:** el texto se descompone en unidades (**tokens**) (sec. 5.2.4), típicamente sub-palabras mediante métodos como BPE o unigram LM; esto permite vocabularios compactos y manejar palabras raras o nuevas.

- **Pre-entrenamiento:** el modelo ajusta millones/miles de millones de parámetros para minimizar el error al predecir el siguiente token.
- **Decodificación:** en uso, el LLM genera texto eligiendo tokens con estrategias como nucleus (top-p) sampling o top-k, que equilibran fluidez y diversidad y evitan salidas repetitivas (degeneración).
- **Adaptación y alineación:** para que el modelo siga instrucciones humanas de forma útil y segura, se emplea (**aprendizaje supervisado**) (sec. 5.3.4): primero se recopilan demostraciones y preferencias humanas, y luego se afinan los pesos del modelo para alinear su conducta con la intención del usuario.

La importancia de características (o feature importance)

La importancia de características (o feature importance) es un conjunto de técnicas que cuantifican la contribución relativa de cada variable de entrada (característica) en un modelo predictivo de aprendizaje automático, determinando qué tan decisiva es cada característica para la salida o predicción del modelo [21] [22] [23].

5.3. Machine learning (ML): el paradigma del aprendizaje a partir de datos

El Machine Learning (ML) es la disciplina que dota a la Inteligencia Artificial de su capacidad de aprendizaje y adaptación. Para el sistema de recomendación de carreras, el ML no es solo una herramienta, sino el núcleo funcional que permite transformar datos de estudiantes en predicciones significativas. Esta sección desglosa qué es el Machine Learning, cómo operan sus modelos y los distintos tipos de aprendizaje, sentando las bases para comprender la metodología específica que se implementará en el proyecto.

Dentro del vasto campo de la IA, el Machine Learning (ML) o Aprendizaje Automático ha emergido como el enfoque más poderoso y transformador [24]. En lugar de programar explícitamente un conjunto de reglas para realizar una tarea, el ML permite a los sistemas aprender estas reglas directamente a partir de los datos [25].

5.3.1. Definición fundamental

El **Machine Learning** es un subcampo de la inteligencia artificial «dedicado al desarrollo y estudio de algoritmos estadísticos que pueden aprender de los datos y generalizar a datos no vistos, y así realizar tareas sin instrucciones explícitas» [26]. La esencia del ML es la capacidad de un sistema para mejorar autónomamente su rendimiento a través de la experiencia, la cual se materializa en forma de grandes volúmenes de datos [27].

5.3.2. Modelo de machine learning

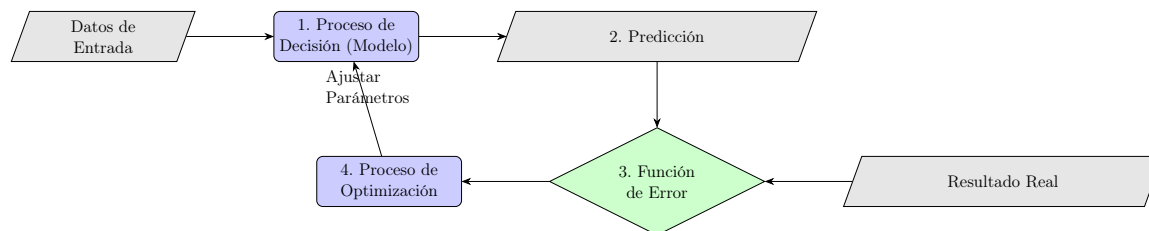
Un **modelo de machine learning** se puede definir como un constructo matemático o programa entrenado a partir de datos, capaz de reconocer patrones y hacer predicciones o tomar decisiones sobre datos nuevos, sin haber sido explícitamente programado para cada escenario concreto [28].

5.3.3. El proceso de aprendizaje

El proceso de aprendizaje en ML puede descomponerse conceptualmente en tres componentes principales: [27]

- **Un proceso de decisión:** el algoritmo se utiliza para hacer una predicción o clasificación. A partir de los datos de entrada, el modelo genera una estimación sobre un patrón en los datos.
- **Una función de error:** esta función evalúa la predicción del modelo. Compara la predicción del modelo con el resultado conocido (en el caso de datos etiquetados) para cuantificar la precisión del modelo. En el protocolo, métricas como la precisión, la sensibilidad y el F1-score cumplen esta función.
- **Un proceso de optimización del modelo:** el algoritmo ajusta iterativamente sus parámetros internos “*pesos*” para minimizar el error medido por la función de error. Este proceso se repite hasta que se alcanza un umbral de precisión deseado.

Figura 1. Diagrama del ciclo de aprendizaje iterativo de un modelo de Machine Learning



5.3.4. Tipos de aprendizaje

Existen varios paradigmas de aprendizaje automático, cada uno adecuado para diferentes tipos de problemas y datos. Los tres principales son [29]:

- **Aprendizaje supervisado (supervised learning):** en este paradigma, el algoritmo se entrena con un conjunto de datos etiquetado, donde cada ejemplo de entrada está emparejado con una etiqueta de salida correcta. El objetivo es aprender una función que mapee las entradas a las salidas para poder predecir la salida de nuevos datos no vistos. «El aprendizaje supervisado utiliza conjuntos de datos de entrada y salida etiquetados por humanos para entrenar modelos de inteligencia artificial» [27].

- **Aprendizaje no supervisado (unsupervised learning):** aquí, el algoritmo trabaja con datos no etiquetados y debe descubrir patrones, estructuras o agrupaciones (clusters) por sí mismo, sin una guía explícita sobre las salidas correctas. Se utiliza para tareas como la segmentación de clientes o el análisis exploratorio de datos.
- **Aprendizaje por refuerzo (reinforcement learning):** en este enfoque, un agente aprende a tomar decisiones interactuando con un entorno. El agente recibe recompensas o castigos (refuerzos) por sus acciones, y su objetivo es aprender una política de acción que maximice la recompensa acumulada a lo largo del tiempo. Es el paradigma utilizado para entrenar IA para jugar juegos o controlar robots.

5.3.5. Entropía cruzada (función de pérdida)

La **entropía cruzada** es una función de pérdida ampliamente utilizada para problemas de clasificación. Mide la discrepancia entre la distribución verdadera de etiquetas y la distribución de probabilidad predicha por el modelo. En clasificación multiclase, la *entropía cruzada categórica* penaliza con mayor fuerza las predicciones altamente seguras pero incorrectas, favoreciendo distribuciones bien calibradas [30], [31].

5.3.6. Hiperparámetros

Los **hiperparámetros** son configuraciones externas al proceso de entrenamiento (p. ej., tasa de aprendizaje, profundidad máxima del árbol, número de iteraciones) que controlan la capacidad y el comportamiento del modelo. A diferencia de los parámetros, no se aprenden directamente de los datos y deben ajustarse mediante validación [30], [32].

5.3.7. Validación cruzada

La **validación cruzada** (p. ej., k-fold) estima el desempeño fuera de muestra dividiendo los datos en k particiones: se entrena en k-1 y se valida en la restante, repitiendo el proceso k veces. Esta técnica reduce la varianza de la estimación respecto a una sola partición entrenamiento/validación [33].

5.3.8. Búsqueda de hiperparámetros

El ajuste de hiperparámetros puede realizarse mediante **búsqueda en rejilla** (grid search) o **búsqueda aleatoria** (random search). La evidencia empírica muestra que la búsqueda aleatoria explora con mayor eficiencia espacios de hiperparámetros de alta dimensión bajo el mismo presupuesto computacional [34].

5.3.9. Regularización y sobreajuste

La **regularización** agrega términos de penalización (p. ej., L1/Lasso, L2/Ridge) o restricciones estructurales para controlar la complejidad del modelo y mitigar el **sobreajuste**, fenómeno en el cual el modelo memoriza ruido del conjunto de entrenamiento y pierde capacidad de generalización. El balance sesgo-varianza guía la elección de la regularización adecuada [31], [32].

5.4. Análisis y preparación de datos

5.4.1. Limpieza del dataset

La **limpieza de datos** es el proceso de identificar y corregir (o eliminar) errores e inconsistencias de un conjunto de datos para mejorar su calidad [35]. Este proceso es fundamental en cualquier proyecto de machine learning, ya que la calidad de los datos de entrada impacta directamente en el rendimiento del modelo.

5.4.2. Técnicas de imputación

Las **técnicas de imputación** son métodos estadísticos utilizados para reemplazar datos faltantes con valores sustitutos [35]. La elección de la técnica de imputación (e.g., media, mediana, o modelos más sofisticados) depende de la naturaleza de los datos y del mecanismo que causa la ausencia de los mismos.

5.4.3. Valores artificiales

En el contexto de la preparación de datos, los **valores artificiales** o sintéticos son datos generados para aumentar o balancear un dataset. Esto puede ser útil en escenarios con datos escasos o desbalanceados, donde el modelo podría beneficiarse de ejemplos adicionales para aprender patrones de manera más efectiva [35].

5.5. Estadística descriptiva

La estadística descriptiva constituye un conjunto de métodos destinados a resumir, organizar y presentar la información derivada de una población o muestra, facilitando la comprensión de los datos y la identificación de patrones generales [36], [37]. En el contexto de la presente investigación, estas técnicas se emplean como una herramienta de apoyo para analizar los resultados obtenidos en las encuestas y validaciones del sistema de recomendación vocacional.

Unidad de análisis: población y muestra

La **población** es el conjunto total de casos/individuos de interés sobre el que deseamos formular conclusiones o estimar parámetros. La inferencia estadística busca aprender sobre esa población a partir de datos muestrales. [38], [39], [40].

Una **muestra** es un subconjunto de la población del cual se recolectan datos para estimar parámetros poblacionales (p. ej., medias, proporciones) y realizar inferencias. [38], [39], [40].

5.5.1. Variables numéricas

Las **variables numéricas** representan cantidades medibles. Pueden ser discretas (toman valores de un conjunto finito) o continuas (toman cualquier valor dentro de un rango) [41].

5.5.2. Medidas de tendencia central

Las **medidas de tendencia central** describen el centro de una distribución de datos.

- **Media:** el promedio aritmético de los datos [41].
- **Mediana:** el valor central de un conjunto de datos ordenado [41].

5.6. Conceptos clave computacionales

5.6.1. Orquestación centralizada

La **orquestación centralizada** es el proceso mediante el cual un controlador (orquestador) centralizado coordina, gestiona y dirige la interacción de múltiples sistemas, servicios o componentes distribuidos para ejecutar un flujo de trabajo completo, asegurando que cada parte actúe en la secuencia correcta y bajo control común. [42] [43]

5.6.2. Arquitectura orientada a servicios

La **arquitectura orientada a servicios (SOA)** es un estilo arquitectónico de software que organiza los sistemas como un conjunto de servicios autónomos, bien definidos, reutilizables e interoperables, que se comunican entre sí mediante interfaces estándar a través de una red, permitiendo construir aplicaciones compuestas a partir de dichos servicios. [44] [45]

5.6.3. Prompt engineering

La Ingeniería de Prompts (Prompt Engineering) es una disciplina dentro de la inteligencia artificial y el procesamiento del lenguaje natural que se centra en el diseño, formulación

y optimización de instrucciones o entradas (prompts) dirigidas a modelos de lenguaje extensos (Large Language Models, LLMs), con el objetivo de controlar y mejorar la calidad, coherencia y pertinencia de las respuestas generadas. Esta práctica combina principios de ingeniería de software, lingüística computacional y diseño de interacción humano-máquina, buscando estructurar las instrucciones de manera que los modelos interpreten correctamente la intención del usuario y produzcan resultados alineados con tareas específicas. [46] [47] [48] [49]

5.6.4. Ítem

Un **ítem** es la unidad básica de un test o cuestionario: un estímulo o pregunta al que la persona responde (p. ej., seleccionando una opción o escribiendo una respuesta). Los ítems se diseñan para capturar porciones del constructo que se desea medir y, en conjunto, forman escalas [50].

5.6.5. Instrumento de evaluación

Un **instrumento de evaluación**, en el contexto psicométrico, es un procedimiento sistemático para observar la conducta y describirla con la ayuda de escalas numéricas o categorías establecidas [51]. Estos instrumentos, comúnmente denominados tests, están diseñados para medir atributos, rasgos o características latentes de las personas que no son directamente observables.

5.6.6. Perfil psicométrico

Un **perfil psicométrico** es la representación gráfica de las puntuaciones de una persona en las distintas escalas o dimensiones que mide un instrumento de evaluación o un conjunto de ellos. Este perfil permite visualizar de forma rápida los puntos fuertes y débiles del individuo evaluado, facilitando la interpretación de los resultados y la toma de decisiones [51].

5.7. Teorías psicológicas

Para que un sistema de recomendación pueda ofrecer orientación vocacional relevante, debe estar fundamentado en teorías psicológicas sólidas que expliquen cómo las características individuales se relacionan con el éxito y la satisfacción profesional. Esta sección introduce tres marcos teóricos clave: el modelo de personalidad de los **Cinco Grandes (Big Five)**, la teoría de los tipos vocacionales de Holland (**RIASEC**) y el concepto de **Grit**. Estos modelos proporcionan la base para evaluar los rasgos de personalidad, los intereses y la perseverancia del estudiante, variables fundamentales para el algoritmo de recomendación.

5.7.1. Big five personality

El **modelo de los Cinco Grandes (Big Five)** es un enfoque ampliamente aceptado en la psicología de la personalidad. Plantea que existen cinco dimensiones básicas de la personalidad humana **Apertura a la experiencia (Openness)**, **Conciencia (Conscientiousness)**, **Extraversión (Extroversion)**, **Amabilidad (Agreeableness)** y **Neuroticismo (Neuroticism)** que juntas abarcan la mayor parte de las diferencias individuales en la personalidad. Estas cinco dimensiones constituyen un espectro continuo en el que cada individuo se puede ubicar, y en conjunto capturan eficientemente la variabilidad de rasgos de personalidad; por ello, el modelo de los Big Five se considera uno de los más utilizados y validados en la investigación actual sobre personalidad [52].

5.7.2. Teoría de holland (Modelo RIASEC)

La Teoría de John L. Holland, o modelo RIASEC, propone que tanto las personas como los entornos de trabajo pueden clasificarse en seis tipos básicos de personalidad/vocación: **Realista**, **Investigador**, **Artístico**, **Social**, **Emprendedor** y **Convencional**. Holland sugiere que existe un principio de congruencia persona-entorno: las personas tienden a estar más satisfechas y tener mayor éxito en sus estudios y carreras cuando sus intereses y habilidades coinciden con las características de su entorno laboral u ocupacional [53].

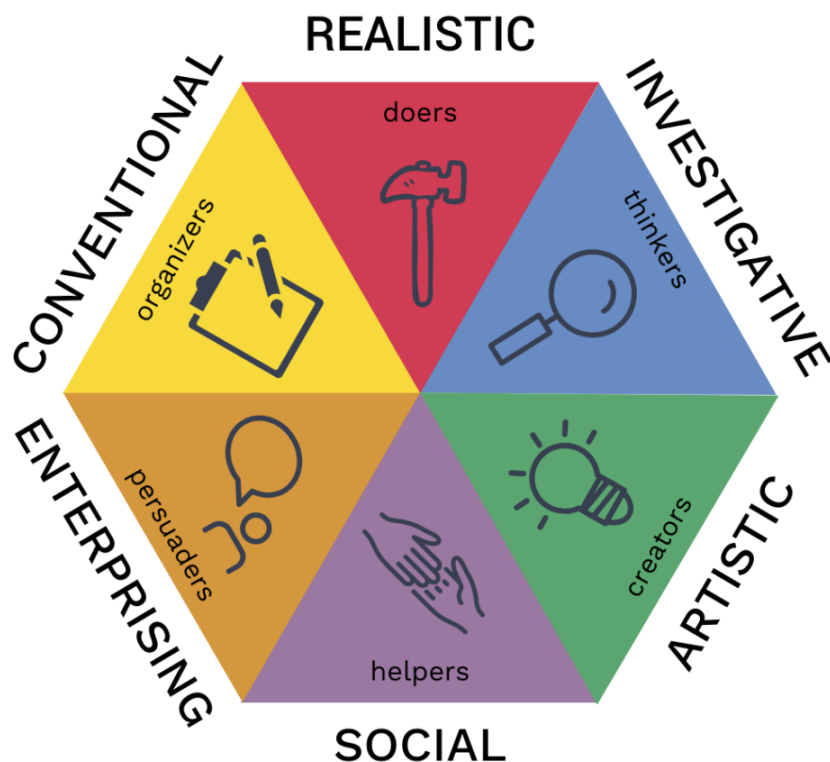
5.7.3. Teoría del grit

La teoría del **"grit"** (determinación) se centra en la importancia de la **perseverancia** y la **pasión** por metas a largo plazo como factores clave del éxito. La idea central del "grit" se puede resumir en lo que eventualmente logramos puede depender más de nuestra pasión y perseverancia que de nuestro talento innato [55]. En este sentido, el "grit" propone que el logro de objetivos difíciles no solo depende de la capacidad o talento, sino también de la constancia del esfuerzo en el tiempo, incluso frente a obstáculos y frustraciones.

5.8. Factores psicosociales en la elección de carrera del estudiante

La elección de una carrera es un proceso moldeado por un complejo entramado de factores psicosociales, como el entorno familiar, el estatus socioeconómico, la presión de grupo y la identidad vocacional. Aunque la dificultad para medir y obtener estos datos de manera sistemática impide su inclusión en nuestro sistema actual, es indispensable analizar su impacto para comprender el panorama completo. Por ello, esta sección explora dichas variables no solo para contextualizar las decisiones del estudiante, sino también para sentar las bases de futuras mejoras que permitan ofrecer una recomendación más holística.

Figura 2. Diagrama del proceso de aprendizaje de un modelo de Machine Learning



Nota. [54].

5.8.1. La influencia del entorno familiar y las expectativas parentales

Las expectativas parentales tienen un papel determinante en la decisión vocacional de los adolescentes, ya que los padres transmiten valores, creencias y aspiraciones que influyen directamente en la percepción de las opciones profesionales. El apoyo y la comunicación familiar fortalecen la autoeficacia académica y, en consecuencia, las decisiones vocacionales [56]. Sin embargo, la presión explícita hacia determinadas carreras puede limitar la autonomía del estudiante, generando conflictos en la construcción de la identidad vocacional [57].

5.8.2. El impacto del estatus socioeconómico en el abanico de opciones percibidas

El nivel socioeconómico condiciona la percepción de oportunidades y acceso a recursos educativos. Jóvenes de contextos desfavorecidos tienden a considerar un rango más limitado de opciones profesionales debido a barreras financieras y culturales [58]. Estudios recientes

enfatan que el capital económico y cultural de la familia es un predictor significativo de la elección de estudios superiores y de la posibilidad de acceder a carreras de alta demanda laboral [59].

5.8.3. La presión de grupo y las tendencias sociales entre pares

La influencia de los pares desempeña un rol visible en la decisión de carrera. La elección no solo responde a intereses individuales, sino también a la necesidad de pertenencia y validación social [60].

5.8.4. El desarrollo de la identidad vocacional durante la adolescencia

La adolescencia es un periodo crítico para el desarrollo de la identidad vocacional. Erikson planteaba que la formación de identidad es central en esta etapa, y estudios posteriores confirman que la indecisión vocacional está ligada al nivel de desarrollo identitario [61].

5.9. Características a medir

Habiendo establecido los fundamentos teóricos, es momento de definir qué características específicas del estudiante se medirán para alimentar el sistema de recomendación. La calidad de las recomendaciones depende directamente de la calidad y pertinencia de los datos de entrada. Esta sección detalla las variables que se recopilarán, abarcando tanto habilidades cognitivas como no cognitivas. La selección de estas características se basa en las teorías psicológicas previamente discutidas y busca crear un perfil integral del estudiante que permita al algoritmo de Machine Learning identificar patrones y realizar predicciones informadas.

Para que el modelo de IA pueda generar recomendaciones personalizadas y significativas, debe ser alimentado con un perfil de usuario multidimensional que capture una imagen holística del individuo. Este perfil se planea construir a partir de tres pilares fundamentales: las habilidades cognitivas, la personalidad, la perseverancia y los intereses vocacionales.

5.9.1. Habilidades cognitivas

Las **habilidades cognitivas** son las capacidades mentales relacionadas con el procesamiento de la información [62]. Son los procesos que permiten la construcción del conocimiento, incluyendo la memoria, la atención, el lenguaje, el razonamiento, la creatividad y la resolución de problemas [63]. Estas habilidades representan el “*hardware*” intelectual del estudiante, su capacidad para adquirir, procesar y utilizar información de manera efectiva. Son fundamentales para el éxito académico y profesional, ya que determinan la facilidad con la que un individuo puede aprender nuevos conceptos y adaptarse a tareas complejas.

5.9.2. Habilidades no cognitivas

Las habilidades no cognitivas (también llamadas soft skills, habilidades socioemocionales o factores no cognitivos) son aquellas actitudes, comportamientos, rasgos de personalidad, motivaciones y estrategias que no se miden mediante pruebas de inteligencia tradicional, pero que resultan fundamentales para lograr éxito académico, profesional y en la vida cotidiana [64].

5.10. Evaluaciones de competencias cognitivas y académicas

Una vez definidas las características a medir, el siguiente paso es seleccionar los instrumentos de evaluación adecuados para recopilar datos fiables y estandarizados. Esta sección describe las pruebas específicas que se utilizarán para medir las competencias cognitivas y académicas de los estudiantes: el PIAAC y la PAA. La elección de estas herramientas no es arbitraria; se basa en su validez psicométrica y su capacidad para proporcionar métricas objetivas del potencial académico del individuo. Estos datos cuantitativos formarán una parte esencial del perfil del estudiante que alimentará al modelo de recomendación.

5.10.1. Programme for the international assessment of adult competencies (PIAAC)

El PIAAC es un estudio comparativo internacional (promovido por la OCDE) que evalúa las competencias cognitivas básicas de la población adulta en tres dominios principales: la comprensión lectora, las matemáticas (numeracy o cálculo) y la resolución de problemas. A través de encuestas y ejercicios prácticos, PIAAC proporciona información detallada sobre el nivel de estas habilidades en los individuos, convirtiéndose en una herramienta valiosa para comprender sus capacidades intelectuales. De hecho, contar con buenas habilidades de lectura y aritmética es fundamental para que las personas adultas puedan aprender nuevas destrezas de forma eficiente y optimizar su potencial de aprendizaje a lo largo de la vida [65].

5.10.2. Prueba de aptitud académica (PAA)

La PAA es un examen estandarizado empleado en varios países hispanohablantes para la admisión en la educación superior. Esta prueba, desarrollada por el College Board, está diseñada para medir las habilidades académicas clave de los estudiantes en áreas como la lectura (razonamiento verbal), la escritura (redacción) y las matemáticas. Los resultados de la PAA sirven como un importante indicador del desempeño potencial de los alumnos en el ámbito universitario; de hecho, muchas instituciones la utilizan para identificar a los aspirantes con mayor probabilidad de éxito en sus estudios superiores [66].

5.11. Contexto profesional

Para que las recomendaciones del sistema sean ancladas en la realidad, es imprescindible definir el "espacio de llegada": el mundo profesional. Esta sección establece el vocabulario clave para entender el objetivo de la orientación vocacional, clarificando los conceptos de profesión, carrera profesional y trayectoria. Comprender estas definiciones es fundamental para contextualizar las metas del sistema de recomendación, que busca no solo sugerir una carrera, sino facilitar el inicio de una trayectoria profesional satisfactoria y coherente con el perfil del individuo.

5.11.1. Profesión

La definición de **profesión** abarca mucho más que simplemente un empleo remunerado. Se trata de una actividad que exige una formación educativa de alto nivel, conocimientos especializados y una capacitación formal, generalmente universitaria. Las profesiones están ligadas al desarrollo social y suelen contar con una organización propia, autorregulación y normas éticas elevadas [67].

En contraste con una ocupación, que se define por las tareas y funciones desempeñadas, la profesión se distingue por su carácter intelectual, la necesidad de una cualificación alta y la existencia de una regulación precisa. En muchos casos, el acceso y ejercicio de una profesión está regulado por requisitos legales y administrativos, especialmente cuando existe un interés público que proteger [67].

5.11.2. Carrera profesional

El concepto de **carrera profesional** es más amplio y dinámico que el de profesión. Se refiere a «una serie de puestos de trabajo que una persona ocupa a lo largo del tiempo» [68]. Implica una trayectoria que se desarrolla a través de la adquisición continua de habilidades, la formación y el establecimiento de metas profesionales. Esta trayectoria no es necesariamente lineal; puede incluir ascensos, cambios laterales entre diferentes roles o sectores, e incluso interrupciones o cambios de dirección. [68]

5.11.3. Trayectoria profesional

La **trayectoria profesional** es un término que enfatiza el recorrido o el camino que un individuo traza a lo largo de su vida laboral. Mientras que "*carrera*" puede connotar una progresión más estructurada, "*trayectoria*" acoge la naturaleza fluida y a veces impredecible del desarrollo profesional en el siglo XXI, que puede implicar múltiples cambios de industria o función. [69]

5.11.4. Orientación vocacional

La **orientación vocacional** es un proceso articulado de acompañamiento psicológico, educativo y social que busca fortalecer el autoconocimiento (intereses, aptitudes, valores) y ofrecer información contextual (sobre formación, mercado laboral y desarrollo profesional), con el propósito de posibilitar decisiones vocacionales conscientes y efectivas, promover un desempeño satisfactorio y facilitar el desarrollo del proyecto de vida y el progreso profesional [70].

5.12. Desafíos de la orientación vocacional en la era digital

La orientación vocacional tradicional se enfrenta a desafíos sin precedentes en el contexto digital actual. Esta sección aborda los problemas clave que el sistema de recomendación propuesto busca mitigar: la paradoja de la información, la rápida evolución del futuro del trabajo y la creciente brecha de habilidades entre la academia y la industria. Comprender estos desafíos es crucial para justificar la necesidad de una herramienta inteligente y personalizada que pueda guiar a los estudiantes a través de la complejidad del panorama educativo y profesional del siglo XXI.

5.12.1. La paradoja de la información

El acceso ilimitado a información a través de internet y redes sociales ha transformado radicalmente la manera en que los estudiantes exploran opciones educativas y profesionales. En teoría, contar con más información debería empoderar al individuo y facilitar una elección más consciente. Sin embargo, múltiples investigaciones muestran que un exceso de alternativas puede tener el efecto contrario: generar confusión, ansiedad e indecisión vocacional. Este fenómeno se conoce como la *paradoja de la información*, en el que el incremento en la cantidad de opciones no se traduce necesariamente en mejores decisiones, sino en saturación cognitiva y desmotivación [71].

La sobreabundancia de datos se hace especialmente evidente en los procesos de orientación vocacional, donde los estudiantes se enfrentan a cientos de carreras, instituciones y posibles trayectorias, cada una acompañada de múltiples fuentes de información (páginas web institucionales, foros, redes sociales, rankings, entre otros). Como señalan [72], este exceso de información puede dar lugar a la llamada “infoxicación”, un estado en el cual la persona percibe que no cuenta con los criterios o herramientas suficientes para evaluar la calidad, la relevancia o la veracidad de la información recibida. En consecuencia, se incrementa la probabilidad de tomar decisiones basadas en factores superficiales o en recomendaciones poco fiables.

En este contexto, los servicios de orientación vocacional digital se enfrentan al reto de no solo proporcionar información, sino también de ****curarla, organizarla y personalizarla****. El objetivo no es maximizar la cantidad de datos disponibles, sino facilitar que el estudiante pueda acceder a información pertinente y alineada con sus intereses, habilidades y valores. Sistemas de recomendación y técnicas de filtrado personalizado representan un avance signi-

ficativo en esta dirección, pues permiten reducir la carga cognitiva al presentar únicamente aquellas alternativas más relevantes para el perfil del usuario [73]. De esta manera, se busca transformar el entorno informativo de uno caótico y desbordante en uno estructurado y comprensible.

Asimismo, la paradoja de la información se ve acentuada por la falta de alfabetización digital en algunos contextos, donde los estudiantes pueden tener dificultades para distinguir entre fuentes confiables y aquellas que difunden información sesgada o incompleta [74]. En consecuencia, además de la personalización tecnológica, resulta esencial fortalecer las competencias de los usuarios en gestión de la información, pensamiento crítico y evaluación de fuentes. La combinación de alfabetización digital con sistemas de recomendación inteligentes puede contribuir a disminuir la incertidumbre y a apoyar decisiones vocacionales más fundamentadas.

5.12.2. El futuro del trabajo y habilidades del siglo XXI

La acelerada transformación tecnológica impulsada por la automatización, la inteligencia artificial, la robótica y la digitalización está redefiniendo de manera profunda el concepto de empleabilidad. El Foro Económico Mundial advierte que cerca del 44 % de las habilidades laborales cambiarán en los próximos cinco años, destacando un desplazamiento hacia competencias que integren tanto capacidades técnicas como socioemocionales [75]. Entre estas, se encuentran la resolución de problemas complejos, el pensamiento crítico, la creatividad, la adaptabilidad y la alfabetización digital.

Investigaciones recientes resaltan que el trabajo del futuro será cada vez más híbrido, requiriendo tanto dominio tecnológico como competencias humanas difíciles de automatizar, como la comunicación efectiva, la empatía y la colaboración interdisciplinaria [76]. En este escenario, los sistemas de orientación vocacional deben dejar de enfocarse únicamente en vincular al estudiante con una profesión estática, para pasar a guiarlo hacia el desarrollo de habilidades transferibles que le permitan navegar trayectorias profesionales dinámicas y en constante cambio [77].

Asimismo, la noción de “aprendizaje a lo largo de la vida” se vuelve esencial. La actualización constante de competencias, a través de programas de reskilling y upskilling, será un requisito para mantener la empleabilidad en un contexto caracterizado por la disrupción tecnológica [78]. Esto implica que la orientación vocacional debe ser entendida como un proceso continuo, y no como un evento único en la adolescencia, alineándose con la necesidad de adaptación permanente que impone el futuro del trabajo.

5.12.3. La brecha de habilidades (skills gap)

La brecha de habilidades constituye uno de los principales retos para los sistemas educativos y laborales contemporáneos. Esta desconexión creciente entre la formación que ofrecen las instituciones educativas y las competencias demandadas por el mercado laboral se manifiesta en múltiples formas: sobrecualificación en ciertos sectores, déficit de talento en áreas tecnológicas emergentes, y dificultades para cubrir puestos especializados [79].

Uno de los factores que alimenta esta brecha es la velocidad con la que evolucionan las industrias frente a la rigidez de los currículos académicos. Mientras que sectores como la inteligencia artificial, la ciberseguridad o las energías renovables avanzan a gran velocidad, los programas formativos muchas veces no logran actualizarse con la misma rapidez [80]. Esto genera un desajuste estructural en el que los egresados carecen de competencias clave al momento de ingresar al mercado laboral.

A nivel social, la brecha de habilidades también contribuye a aumentar desigualdades, ya que los trabajadores de menor nivel educativo suelen tener menos acceso a programas de actualización y, en consecuencia, corren un mayor riesgo de quedar desplazados del mercado laboral [81]. En este sentido, los sistemas de orientación vocacional basados en inteligencia artificial representan una oportunidad para anticipar las tendencias del mercado y alinear la formación académica con la demanda real de competencias, contribuyendo así a reducir los desajustes y favorecer la equidad en el acceso a oportunidades profesionales.

5.13. Modelos y enfoques en la orientación vocacional

A lo largo del tiempo, la orientación vocacional ha evolucionado desde modelos tradicionales centrados en el consejero hasta enfoques tecnológicos emergentes que aprovechan el poder de los datos y la inteligencia artificial. Esta sección traza una línea evolutiva entre ambos paradigmas, destacando las limitaciones de los métodos clásicos y las oportunidades que ofrecen las nuevas tecnologías. Este análisis es fundamental para posicionar el sistema de recomendación propuesto como una solución que no solo se alinea con las tendencias digitales, sino que también responde a las deficiencias de los enfoques tradicionales, ofreciendo una alternativa más escalable, personalizada y adaptada al contexto actual.

5.13.1. Modelos tradicionales

Los enfoques clásicos de orientación vocacional surgieron principalmente a mediados del siglo XX, cuando el acompañamiento al estudiante se estructuraba en torno a la figura del consejero o psicólogo educativo. Estos modelos se basaban en entrevistas personales, administración de pruebas psicométricas en formato papel y sesiones de retroalimentación individual o grupal. La finalidad era ayudar al estudiante a identificar sus intereses, aptitudes y valores, y vincularlos con las opciones educativas y profesionales disponibles en su contexto inmediato.

Aunque estos métodos tuvieron un impacto positivo en el fortalecimiento de la identidad vocacional, presentan limitaciones significativas frente a los retos contemporáneos. En primer lugar, la naturaleza estática de las pruebas aplicadas en papel no permite una actualización constante ni la incorporación de nuevas dimensiones relacionadas con los cambios en el mercado laboral [82]. En segundo lugar, los modelos tradicionales suelen ser intensivos en recursos humanos, lo que dificulta ofrecer atención personalizada a un número amplio de estudiantes, especialmente en instituciones con recursos limitados. Finalmente, este enfoque tiende a ser lineal y unidireccional, centrado en la transmisión de información, sin aprovechar estrategias dinámicas de retroalimentación ni la riqueza de datos que los estudiantes generan

en entornos digitales.

5.13.2. Enfoques tecnológicos emergentes

Con el avance de la digitalización y las tecnologías educativas, han surgido nuevos enfoques que transforman la orientación vocacional en un proceso más interactivo, dinámico y personalizado. Las plataformas en línea permiten integrar diferentes recursos —como cuestionarios automatizados, perfiles de usuario, algoritmos de recomendación y simulaciones de carreras— en un ecosistema accesible desde cualquier dispositivo. Esto ha democratizado el acceso a los servicios de orientación, reduciendo barreras geográficas y temporales.

Uno de los elementos más destacados en estos enfoques es la incorporación de la **gamificación**, que utiliza mecánicas propias de los videojuegos (recompensas, niveles, desafíos) para motivar a los estudiantes a explorar opciones profesionales y reflexionar sobre sus intereses. Diversos estudios han demostrado que la gamificación incrementa la motivación intrínseca y mejora la calidad del aprendizaje vocacional [83].

Asimismo, el uso de **inteligencia artificial** y **análisis de datos** ha abierto nuevas posibilidades para la personalización. Por ejemplo, algoritmos de machine learning pueden analizar patrones en las respuestas de pruebas psicométricas, historial académico o intereses declarados para recomendar trayectorias educativas más acordes al perfil del estudiante. Estas recomendaciones no solo reducen la sobrecarga de información, sino que también generan itinerarios dinámicos que se adaptan conforme el usuario interactúa con la plataforma [84].

En síntesis, mientras los modelos tradicionales sentaron las bases de la orientación vocacional como práctica educativa, los enfoques tecnológicos emergentes ofrecen herramientas más flexibles, escalables y adaptadas a la complejidad del siglo XXI. Ambos enfoques no deben entenderse como opuestos, sino como complementarios, donde la tecnología potencia y amplía el alcance de los fundamentos clásicos de la orientación vocacional.

5.14. Estado del arte de los sistemas de recomendación en contextos educativos (EdRecSys)

Los sistemas de recomendación han permeado diversas industrias, y el sector educativo no es la excepción. Esta sección revisa el estado del arte de los Sistemas de Recomendación en Contextos Educativos (EdRecSys), explorando cómo se han utilizado para sugerir desde material de estudio hasta carreras profesionales. Analizar las implementaciones existentes es crucial para identificar tanto las mejores prácticas como las brechas tecnológicas. Este panorama permite contextualizar la contribución del presente proyecto, mostrando cómo se apoya en los avances existentes y, a su vez, cómo busca superar las limitaciones de los sistemas actuales.

5.14.1. Recomendación de material

Plataformas educativas masivas como Coursera, edX o UdeMY han incorporado algoritmos de recomendación que utilizan tanto el historial de aprendizaje del usuario como las preferencias declaradas para sugerir cursos relevantes. Estos sistemas, generalmente basados en filtrado colaborativo o filtrado basado en contenido, buscan mejorar la experiencia del estudiante al reducir la sobrecarga de información y facilitar el descubrimiento de contenidos pertinentes. De acuerdo con [85], los sistemas de recomendación en contextos de aprendizaje mejoran la personalización de la experiencia educativa y aumentan las tasas de finalización de cursos, ya que logran adaptar el recorrido de aprendizaje a los intereses y necesidades individuales. Además, investigaciones recientes muestran que la efectividad de estos sistemas radica en su capacidad de integrar múltiples fuentes de datos —como tiempo de dedicación, niveles de interacción y desempeño en evaluaciones— para refinar continuamente las sugerencias [86].

5.14.2. Recomendación de carreras profesionales

Si bien la aplicación de sistemas de recomendación a la orientación vocacional es todavía incipiente, en los últimos años se han desarrollado propuestas que utilizan perfiles psicométricos y datos académicos para vincular a los estudiantes con posibles trayectorias profesionales. Investigaciones como la de [87] destacan el potencial de aplicar algoritmos de machine learning y técnicas de minería de datos para identificar patrones entre características individuales —como intereses, habilidades y valores— y la elección de carrera. Estos sistemas buscan superar las limitaciones de los enfoques tradicionales, ofreciendo una orientación dinámica y personalizada que considera tanto las preferencias del estudiante como la evolución del mercado laboral. A diferencia de los modelos de recomendación de cursos, la recomendación de carreras implica un grado mayor de complejidad, ya que requiere integrar factores psicosociales, datos académicos históricos y predicciones sobre las competencias que serán demandadas en el futuro [79]. Aunque los resultados preliminares son prometedores, la literatura enfatiza que aún se deben superar retos como la escasez de bases de datos integrales, la necesidad de modelos explicables y la garantía de equidad en las recomendaciones para evitar sesgos discriminatorios.

5.15. El rol de los modelos de lenguaje extensos (LLMs) en la guianza personalizada

Los Modelos de Lenguaje Extensos (LLMs) representan la vanguardia de la Inteligencia Artificial y ofrecen oportunidades sin precedentes para personalizar la interacción en sistemas de recomendación. Esta sección explora cómo los LLMs pueden trascender la mera presentación de resultados para actuar como agentes conversacionales, tutores o consejeros vocacionales. Se analiza su capacidad para personalizar el contenido y su integración con bases de datos específicas mediante técnicas como RAG (Retrieval-Augmented Generation). Comprender el rol de los LLMs es fundamental, ya que constituyen una de las tecnologías centrales propuestas en este proyecto para crear una experiencia de usuario más natural,

interactiva y efectiva.

5.15.1. IA conversacional como tutor o consejero

En los últimos años, los agentes conversacionales se han consolidado como una herramienta prometedora en el ámbito educativo, al facilitar la interacción natural entre estudiantes y sistemas de información. Los LLMs (Large Language Models), gracias a su capacidad de procesar lenguaje natural y generar respuestas coherentes, ofrecen un potencial significativo para simular el rol de tutores o consejeros vocacionales. [88] destacan que estos sistemas no solo pueden proporcionar información, sino también actuar como acompañantes en el proceso de toma de decisiones, incrementando la motivación y el sentido de apoyo percibido por los estudiantes.

Además, los chatbots basados en IA permiten una disponibilidad constante, superando las limitaciones de tiempo y espacio que presentan los modelos tradicionales de orientación vocacional. Estudios recientes muestran que los estudiantes tienden a interactuar con estos sistemas de manera más abierta que con consejeros humanos, ya que perciben menor juicio social, lo cual facilita la autoexploración de intereses y opciones profesionales [89]. Así, la IA conversacional no solo complementa el trabajo de los orientadores humanos, sino que abre nuevas oportunidades para llegar a poblaciones que históricamente han tenido un acceso limitado a estos servicios.

5.15.2. Personalización del contenido

Una de las principales ventajas de los LLMs en orientación vocacional es su capacidad de personalizar la información en función del perfil del estudiante. A diferencia de los recursos estáticos, estos modelos pueden adaptar el nivel de detalle, la complejidad del lenguaje y los ejemplos utilizados a las características cognitivas, intereses y nivel académico del usuario. Por ejemplo, la descripción de una carrera universitaria en ingeniería puede explicarse de manera distinta a un estudiante de secundaria que apenas inicia en ciencias, en comparación con alguien que ya ha cursado asignaturas de matemáticas avanzadas.

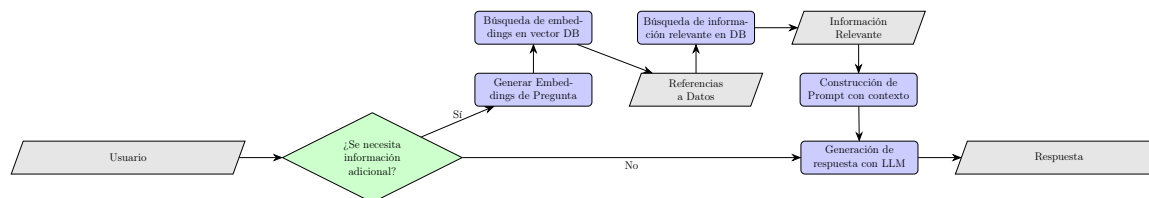
[90] subrayan que los LLMs tienen la capacidad de transformar contenidos densos en explicaciones accesibles y contextualizadas, facilitando así la comprensión de planes de estudio, requisitos académicos y trayectorias profesionales. Esta adaptación dinámica del contenido también permite atender a estilos de aprendizaje diversos (visual, verbal, práctico), aumentando la probabilidad de que el estudiante incorpore la información en su proceso de toma de decisiones.

De igual forma, los modelos pueden integrar información contextual actualizada, como tendencias del mercado laboral, salarios promedio o la demanda proyectada de ciertas profesiones, generando explicaciones que no solo son personalizadas sino también pertinentes al entorno socioeconómico del usuario. Esto representa un salto cualitativo frente a los enfoques tradicionales de orientación, donde la información tiende a ser más homogénea y menos contextualizada.

5.15.3. Retrieval-augmented generation (RAG)

Retrieval-Augmented Generation (RAG) es un enfoque que combina la capacidad generativa de un modelo de lenguaje extenso (LLM) con un componente de recuperación de información. En lugar de depender únicamente del conocimiento interno del modelo, RAG le permite consultar fuentes de datos externas y autorizadas en tiempo real para obtener información relevante y actualizada. Este proceso de dos pasos —recuperar y luego generar— permite al modelo basar sus respuestas en evidencia concreta y verificable, reduciendo la probabilidad de generar información incorrecta (alucinaciones) y permitiendo que sus conocimientos sean fácilmente actualizados [91].

Figura 3. Diagrama de Retrieval-Augmented Generation (RAG)



5.15.4. Conectar LLMs con bases de datos específicas y acceso a información privada

Una de las aplicaciones más poderosas de RAG es la capacidad de conectar un LLM a bases de datos privadas, como los resultados psicométricos de un estudiante o los catálogos de carreras de una institución. En lugar de reentrenar costosamente el modelo con esta información, RAG permite que los datos relevantes sean recuperados y proporcionados al LLM como contexto para generar una respuesta. En el contexto de orientación vocacional, esto permite que el agente conversacional consulte catálogos institucionales de carreras, descripciones oficiales de planes de estudio, resultados psicométricos del propio estudiante y estadísticas laborales locales sin necesidad de modificar el modelo base.

Riesgos asociados a la conexión con bases de datos privadas

Conectar un LLM a datos privados introduce riesgos técnicos, legales y éticos que deben abordarse explícitamente:

- **Fuga de información (data leakage).** fragmentos sensibles (p. ej. resultados psicométricos, identificadores personales, observaciones clínicas) podrían incorporarse inadvertidamente en las respuestas si no se filtran o redaccionan adecuadamente.
- **Memorización del modelo.** aunque la información se recupera en cada consulta, ciertos pipelines mal diseñados pueden acabar almacenando datos sensibles en cachés o en logs accesibles, o incluso inyectándolos en retrainings posteriores.

- **Acceso no autorizado.** si el índice o la base de datos están mal protegidos (endpoints públicos, credenciales débiles), un atacante podría consultar datos privados.
- **Información desactualizada o errónea.** si las fuentes no se versionan ni se actualizan, el sistema podría devolver datos obsoletos (planes de estudio cambiados, vacantes laborales).
- **Sesgos y discriminación.** la evidencia recuperada puede reflejar desigualdades históricas (p. ej. salarios por género) y el LLM puede reforzar narrativas sesgadas si no hay mecanismos de mitigación.
- **Trazabilidad insuficiente.** sin metadatos de procedencia, es difícil explicar por qué el sistema recomendó cierta carrera, lo que afecta la confianza y la responsabilidad.

Medidas técnicas y de gobernanza recomendadas

A continuación se proponen controles técnicos, operativos y de gobernanza que deberías incorporar en la metodología y en la implementación:

Controles de acceso y arquitectura

- **Segmentación de red y despliegue en VPC:** alojar índices y servicios de recuperación en redes privadas (VPC) y habilitar endpoints privados para la comunicación con el LLM.
- **Autenticación y RBAC:** implementar autenticación fuerte (OAuth2 / mTLS) y control de acceso por roles para limitar quién/qué servicio puede consultar datos sensibles.
- **Cifrado:** cifrar los datos en reposo y en tránsito (TLS 1.2/1.3). Para índices vectoriales, apoyar cifrado a nivel de disco y, si es posible, cifrado por vectores/columnas.

Control de la calidad de la evidencia y prevención de “alucinaciones”

- **Verificación y puntuación de fragmentos:** el retriever debe devolver fragmentos con un score y metadatos (fuente, fecha, versión). El generador debe citar explícitamente las fuentes usadas en la respuesta.
- **Sanitización de respuestas:** implementar reglas que detecten y bloqueen la divulgación de PII en la output del LLM (filtros de redacción automática).

Seguridad operativa y pruebas

- **Pruebas de penetración y revisión de seguridad:** auditorías periódicas de la infraestructura y del índice; pruebas de inyección adversarial en prompts y queries.

- **Monitorización y detección de anomalías:** alertas sobre patrones de acceso inusuales, tasas de queries elevadas o intentos de extracción masiva (data exfiltration).
- **Políticas de backups y borrado seguro:** asegurar que los backups también respeten cifrado y retención; procedimientos para borrado seguro y verificación de eliminación.
- **Implementación de rol de lectura:** utilizar un rol de lectura para la extracción de datos por parte del modelo, para evitar eliminación, agregación o alteración de datos por parte del modelo.

5.16. Componentes específicos del modelo propuesto

Habiendo explorado los conceptos teóricos de alto nivel, esta sección desciende al plano técnico para definir los componentes específicos que conformarán la arquitectura del sistema de recomendación. Se detallan conceptos clave como motor de recomendación, conjunto de datos (dataset), características (features) y métricas de evaluación. Comprender estos elementos es esencial para entender la implementación práctica del proyecto, ya que describen las piezas fundamentales sobre las que se construirá y evaluará el modelo de Machine Learning.

5.16.1. Motor de recomendación

Un **motor de recomendación**, o sistema de recomendación (RecSys), es una subclase de sistema de filtrado de información que proporciona sugerencias de ítems que son más pertinentes para un usuario en particular. Estos sistemas son cruciales en entornos con una sobrecarga de opciones, como el comercio electrónico, el streaming de contenido y, en este caso, la elección de una carrera profesional. Existen dos enfoques principales para construir estos modelos: [92]

5.16.2. Tipos de motores de recomendación

Los motores de recomendación se pueden clasificar en varias categorías según su enfoque y técnica subyacente. se definen las siguientes tres categorías principales [92]:

1. **Filtrado colaborativo:** este sistema ofrece recomendaciones basándose en la similitud entre las preferencias de diferentes usuarios. Asume que si a una persona le gusta algo, a usuarios con gustos parecidos también les gustará. Amazon y Spotify son ejemplos de empresas que utilizan este método. Su principal debilidad es el “*problema del arranque en frío*”, que ocurre cuando hay pocos datos sobre nuevos usuarios. Existen dos tipos principales:
 - **Basados en memoria:** utilizan una matriz de usuarios y artículos para encontrar “*vecinos*” o similitudes, ya sea comparando usuarios entre sí (basado en el usuario) o artículos según la interacción de los usuarios con ellos (basado en elementos).

- **Basados en modelos:** crean un modelo de machine learning para predecir qué artículos le podrían gustar a un usuario. Una técnica común es la factorización matricial, que descompone la matriz de datos para predecir las valoraciones faltantes.
2. **Filtrado basado en contenido:** este sistema recomienda artículos basándose en sus propias características (como color, categoría o precio). La idea principal es que si te gustó un producto, te gustarán otros con atributos similares. A diferencia del filtrado colaborativo, no sufre tanto el problema del “*arranque en frío*”, ya que se basa en las descripciones de los artículos y no en el historial del usuario. Sin embargo, su principal limitación es que tiende a recomendar cosas muy parecidas a las que ya conoces, dificultando el descubrimiento de nuevos intereses.
 3. **Sistema de recomendación híbrido:** como su nombre lo indica, este sistema combina el filtrado colaborativo y el filtrado basado en contenido para aprovechar las fortalezas de ambos. Este enfoque puede mejorar significativamente la calidad de las recomendaciones, aunque requiere una arquitectura más compleja y mayor capacidad de cómputo. Un ejemplo claro de su uso es Netflix, que lo emplea para sugerir películas y series.

5.16.3. Conjunto de datos (Dataset)

Un **conjunto de datos** es una «colección estructurada de datos, organizados y almacenados juntos para su análisis o procesamiento» [93]. Es la materia prima para cualquier proyecto de machine learning. La calidad, relevancia y representatividad del conjunto de datos son determinantes para el éxito del modelo [94].

5.16.4. División del dataset

- **Dataset de entrenamiento:** es el subconjunto específico del dataset total que se utiliza para entrenar el modelo de machine learning [95]. Es en esta porción de los datos donde el algoritmo busca y aprende los patrones.
- **Dataset de validación:** es una muestra de datos retenidos del entrenamiento del modelo para estimar la habilidad del modelo mientras se ajustan los hiperparámetros» [96].
- **Dataset de prueba:** es un conjunto de datos sobre el que vamos a aplicar nuestro modelo una vez ha sido entrenado para confirmar su rendimiento real» [97].

5.16.5. Feature (característica)

En el contexto del machine learning, una **feature** (característica/atributo) es una «propiedad o característica individual medible de un conjunto de datos» [98]. Las features son las variables de entrada que el modelo utiliza para hacer predicciones. La selección y construcción de features informativas, discriminatorias e independientes es un paso crucial conocido como ingeniería de características (feature engineering) [98].

5.16.6. Feature engineering

La **ingeniería de características** (feature engineering) es el proceso de transformar data cruda en información relevante para utilizarla en el entrenamiento de un modelo de Machine Learning [99]. Este proceso puede incluir:

- **Selección de características:** identificar y seleccionar las variables más relevantes para el modelo, eliminando las que no aportan información útil.
- **Transformación de características:** modificar las variables existentes para que sean más adecuadas para el modelo, como normalizarlas, estandarizarlas o aplicarles funciones matemáticas.
- **Creación de nuevas características:** generar nuevas variables a partir de las existentes, combinando o descomponiendo información para capturar patrones más complejos.

5.16.7. Métricas de evaluación

Las métricas de evaluación se utilizan para monitorear el rendimiento de modelos de Inteligencia Artificial, proporcionando información a lo largo del ciclo de vida del modelo [100].

Tipos de métricas de evaluación

Las métricas de evaluación se pueden clasificar en diferentes tipos según su propósito y la etapa del ciclo de vida del modelo en la que se aplican. Algunas de las categorías más comunes incluyen [100]:

- **Métricas de deriva:** ayuda a determinar caídas en la precisión y coherencia de los datos en nuestros modelos, para determinar la efectividad del modelo a lo largo del tiempo.
- **Métricas de equidad:** ayudan a determinar si nuestro modelo produce resultados sesgados. En otras palabras, podemos identificar si nuestro modelo muestra una tendencia a proporcionar resultados favorables a cierto grupo sobre los otros.
- **Métricas de vigilancia de la salud del modelo:** monitorea el estado del modelo, proporcionando información sobre el comportamiento y la rendimiento del mismo, para determinar la eficacia con la que el despliegue del modelo procesa las transacciones.
- **Métricas de evaluación de calidad:** permiten medir la capacidad del modelo para proporcionar resultados precisos y útiles, asegurando que cumpla con los estándares requeridos.

Métricas de ranking: DCG/NDCG

En tareas de recomendación y *learning-to-rank*, la calidad del **orden** es tan importante como la clasificación. La **Ganancia Acumulada con Descuento** (DCG) evalúa la utilidad de los ítems relevantes ponderando su posición en la lista; su versión normalizada **NDCG** facilita comparaciones entre consultas al escalar por el DCG ideal. Es especialmente útil para medir la calidad de las primeras posiciones (p. ej., NDCG@5) [101].

5.16.8. Learning-to-Rank (LTR)

Learning-to-Rank aborda el problema de ordenar ítems por relevancia. Existen enfoques *pointwise*, *pairwise* y *listwise*; métodos como **LambdaMART** optimizan funciones de ranking mediante gradientes diseñados para mejorar métricas como NDCG, y han sido ampliamente adoptados en sistemas de recomendación y búsqueda [102], [103].

5.17. Herramientas utilizadas

La materialización de un sistema de recomendación basado en Machine Learning requiere de un conjunto de herramientas tecnológicas específicas. Esta sección describe el stack tecnológico empleado en el proyecto, detallando el rol de cada componente. Se justifica la elección de Python como lenguaje de programación, junto con librerías clave como Pandas, Scikit-learn y XGBoost para la manipulación de datos y la construcción de modelos. Además, se menciona el uso de PsyToolkit para la recolección de datos psicométricos. Describir estas herramientas es fundamental para asegurar la replicabilidad del estudio y para contextualizar las decisiones de implementación.

5.17.1. Python

Python es un lenguaje de programación de alto nivel, interpretado y de propósito general; su legibilidad y sintaxis clara han impulsado su adopción en ciencia de datos y aprendizaje automático [104]. Es conocido por su ecosistema científico (por ejemplo, NumPy y SciPy) que provee bloques fundamentales para el cómputo científico y análisis numérico, lo que explica su popularidad en investigación y aplicaciones de IA/ML [105]. En el contexto de este estudio, Python se utilizó para el análisis de datos y la implementación de modelos de aprendizaje automático.

5.17.2. Librería pandas

Pandas es una biblioteca de Python especializada en la manipulación y análisis de datos que ofrece estructuras de datos como DataFrame y Series para trabajar con datos de forma eficiente e intuitiva; el ecosistema científico reciente muestra que el API tipo DataFrame de Pandas es un estándar de facto para preparación y análisis de datos [106]. Además, numerosas herramientas científicas se construyen directamente sobre DataFrames de Pandas,

lo que evidencia su papel central en la analítica [107]. En este estudio, Pandas se utilizó para la limpieza, transformación y análisis exploratorio de los datos, facilitando la identificación de patrones y tendencias.

5.17.3. Librería scikit-learn

Scikit-learn es una biblioteca de Python diseñada para aprendizaje automático, que proporciona herramientas simples y eficientes para minería y análisis de datos (p. ej., clasificación, regresión, clustering) y flujos de trabajo con pipelines [104]. La literatura reciente subraya que bibliotecas fáciles de usar como scikit-learn han acelerado la adopción de flujos de trabajo de ML en investigación y aplicaciones, gracias a su API unificada de estimadores (fit/predict) y pipelines [108]. En este estudio, scikit-learn se utilizó para implementar y evaluar modelos de aprendizaje automático, permitiendo identificar relaciones complejas y realizar predicciones.

5.17.4. Psytoolkit

PsyToolkit es una plataforma de código abierto para crear y ejecutar experimentos psicológicos y encuestas en línea; la publicación original describe que permite programar experimentos de tiempo de reacción y embebidos en cuestionarios con herramientas de diseño, recolección y análisis de datos [109]. Estudios comparativos han mostrado que PsyToolkit es viable para experimentos web y puede replicar efectos de laboratorio en tareas de respuesta y tiempo de reacción, respaldando su uso en investigación en psicología y ciencias cognitivas [110].

5.17.5. XGBoost

XGBoost (eXtreme Gradient Boosting) es una biblioteca de aprendizaje automático de código abierto que proporciona un marco de trabajo para el aumento de gradiente. Es ampliamente reconocida por su rendimiento y velocidad, lo que la ha convertido en una herramienta popular para problemas de clasificación, regresión y clasificación [111]. En comparación con otros marcos de aumento de gradiente como LightGBM, XGBoost utiliza una estrategia de crecimiento a nivel que puede conducir a una mejor precisión, aunque a veces a costa de tiempos de entrenamiento más largos. Para este proyecto, se eligió XGBoost por su robustez y su capacidad para manejar datos dispersos de manera eficiente, lo cual es crucial para el sistema de recomendación de carreras. La elección de XGBoost fue un pilar en el desarrollo de este trabajo, ya que su algoritmo ‘XGBRanker’ se utilizó para entrenar el modelo central de aprendizaje para clasificación (*learning-to-rank*).

5.17.6. Vector DB

Una **base de datos vectorial (VDB)** administra y consulta representaciones numéricas de alto dimensionalidad (**embeddings**) (sec. 5.2.4) con operaciones de similitud

(ANN, k-NN) y consultas híbridas (filtros por atributos + similitud). Los retos de diseño incluyen: ambigüedad de la similitud semántica, tamaño de los vectores, costo de cómputo de comparaciones, particionado/índices y consultas híbridas eficientes. Los VDB modernos han surgido por demandas de LLMs y búsquedas semánticas a gran escala, sobre la base de décadas de trabajo en búsqueda por similitud [112].

Dimensionalidad de los embeddings

Es el número de componentes de cada embedding almacenado/consultado en la Vector DB. La dimensionalidad viene dada por el modelo que genera los embeddings (p. ej., 384, 768, 1024) y afecta directamente memoria, latencia y recall: a mayor dimensionalidad, más coste de cómputo y RAM por vector/consulta.

Tipo de vector de los embeddings

Describe las propiedades representacionales y de almacenamiento de los embeddings. Ejemplos: Densos, Esparsos.

Función de los embeddings

Es la regla con la que la VDB ordena vecinos cercanos, la métrica elegida determina el índice y el ranking.

Pinecone

Pinecone es un servicio gestionado y “serverless” de base de datos vectorial en la nube que abstrae la elección de algoritmos e índices, ofreciendo escalado automático, latencias bajas y relevancia consistente para búsquedas semánticas y RAG en producción. Su arquitectura separa plano de consulta y de almacenamiento para alta disponibilidad y rendimiento estable a escala [113] [114].

5.17.7. Bases de datos no SQL

“NoSQL” (not only SQL) agrupa motores no relacionales con modelos de datos variados (clave-valor, documento, columna ancha, grafos). En entornos distribuidos suelen balancear propiedades del teorema CAP (consistencia, disponibilidad, tolerancia a particiones) según su caso de uso, por ejemplo priorizando disponibilidad y particiones con consistencia eventual. Estas bases suelen ofrecer esquemas flexibles y escalado horizontal sencillo [115].

Bases de datos documental

Un sistema documental almacena datos como documentos (JSON/BSON) con estructura jerárquica y flexible; permite evolucionar el esquema y modelar objetos de aplicación de forma natural, además de escalar horizontalmente. MongoDB es un ejemplo representativo: persiste documentos en BSON (superset binario de JSON) y expone consultas ricas sobre dichos documentos [116] [117].

DocumentDB

En la práctica académica/industrial, “DocumentDB” suele referirse a Amazon DocumentDB (con compatibilidad MongoDB), un servicio administrado para cargas documentales JSON que mantiene réplicas en múltiples zonas de disponibilidad y automatiza detección/recuperación de fallos a nivel de volumen de clúster, con configuraciones de almacenamiento estándar u optimizadas por I/O. (Históricamente, Microsoft usó “Azure DocumentDB”, hoy Cosmos DB, pero aquí nos referimos al servicio de AWS) [118].

Postman

La herramienta *Postman* es una plataforma todo-en-uno para construir, probar, colaborar y monitorizar APIs, que facilita cada etapa del ciclo de vida de la API: diseño, pruebas, entrega y supervisión [119].

Collection runner La función *Collection Runner* de Postman permite ejecutar un conjunto (colección) de solicitudes API guardadas en un orden definido, registrar los resultados de las pruebas para cada solicitud y automatizar flujos de trabajo entre peticiones [120].

5.18. Consideraciones éticas y desafíos de la IA en la orientación vocacional

El uso de Inteligencia Artificial en un ámbito tan sensible como la orientación vocacional conlleva una gran responsabilidad. No basta con desarrollar un modelo técnicamente robusto; es imperativo abordar sus implicaciones éticas. Esta sección final del marco teórico se dedica a analizar los desafíos críticos asociados a la implementación de IA en este contexto: los sesgos algorítmicos, la privacidad de los datos del estudiante y el problema de la caja negra.^o falta de interpretabilidad. Reflexionar sobre estos temas es crucial para diseñar un sistema que no solo sea efectivo, sino también justo, transparente y respetuoso con el usuario.

5.18.1. Sesgos algorítmicos

Uno de los riesgos más significativos de la IA aplicada a la orientación vocacional es la reproducción y amplificación de sesgos sistémicos de género, etnia o clase social [121]. Si los datos de entrenamiento reflejan desigualdades históricas del mercado laboral, por ejemplo, una menor representación de mujeres en carreras STEM o de ciertos grupos minoritarios en posiciones de liderazgo, el modelo podría estar sesgado en este campo. Como resultado, el sistema podría desaconsejar sistemáticamente ciertas trayectorias profesionales a grupos subrepresentados, no por falta de aptitud, sino por una falta de representación, limitando las oportunidades de los estudiantes.

5.18.2. Privacidad de los datos del estudiante

La recopilación de datos psicométricos y personales para alimentar el modelo plantea desafíos éticos cruciales. La información sobre las habilidades cognitivas, la personalidad, los intereses y las aspiraciones de un estudiante es intrínsecamente sensible. Por ello, es imperativo adherirse a principios éticos estrictos, como el consentimiento informado, donde los usuarios comprendan claramente qué datos se recopilan, cómo se utilizarán y con qué fines. Garantizar la anonimización, el almacenamiento seguro y el uso exclusivo de los datos para el propósito de la orientación es esencial no solo para cumplir con las regulaciones de privacidad, sino para construir y mantener la confianza del estudiante en el sistema [122].

5.18.3. El problema de la caja negra (black box)

La interpretabilidad de los modelos complejos de aprendizaje automático, a menudo denominados “cajas negras”, sigue siendo un reto fundamental, especialmente en aplicaciones de alto impacto como la orientación vocacional [123]. Una recomendación de carrera sin una justificación clara puede generar desconfianza y socavar la autonomía del estudiante, quien necesita comprender el ‘porqué’ detrás de una sugerencia para poder tomar una decisión informada. Por ello, es crucial incorporar técnicas de Inteligencia Artificial Explicable (XAI) que puedan traducir las predicciones del modelo en explicaciones comprensibles. En este proyecto, por ejemplo, se analizan las contribuciones de las características para que el estudiante pueda ver qué factores de su perfil (como la afinidad de intereses o las habilidades cognitivas) tuvieron mayor peso en una recomendación específica, fomentando así la transparencia y la confianza.

Inteligencia artificial explicable (XAI)

La Inteligencia Artificial Explicable (XAI) es un campo de la inteligencia artificial que desarrolla técnicas, procesos y metodologías con el fin de generar explicaciones comprensibles para humanos acerca de las predicciones, decisiones o recomendaciones realizadas por sistemas automatizados complejos (por ejemplo, modelos de aprendizaje automático “caja negra”). Estas explicaciones buscan aumentar la transparencia, la confianza, la responsabilidad y la supervisión sobre dichos sistemas [124] [125] [126] [127].

La investigación es de tipo aplicado, experimental y tecnológico, ya que integra la implementación de diversas tecnologías para el desarrollo de un sistema de orientación vocacional. Para su construcción, se adopta una metodología ágil, la cual permite un proceso iterativo y adaptable durante el desarrollo. Esta metodología facilita abordar de manera estructurada los principales componentes del sistema: la gestión de la información utilizada como base de conocimiento, la guianza vocacional orientada a responder preguntas de los usuarios, la recomendación de carreras y generación de contenido temático personalizado, así como la comunicación eficiente entre el sistema y el usuario.

6.1. Diseño del sistema

El sistema debe de ser capaz de recomendar carreras universitarias y contenido, y, a su vez, debe de poder responder preguntas relacionadas con carreras universitarias, planes de estudio y contenidos temáticos, utilizando información confiable y actualizada proveniente de una fuente con datos institucionales.

6.1.1. Justificación del desacoplamiento

El desacoplamiento del sistema es necesario para poder escalar el sistema de manera independiente y para poder mantener el sistema de manera modular.

Beneficios del desacoplamiento:

- **Reutilización:** permite reutilizar los módulos en otros sistemas.
- **Mantenimiento:** permite realizar cambios en cada módulo de manera independiente, sin afectar los otros módulos.

- **Escalado:** permite escalar cada módulo de manera independiente.
- **Despliegue:** permite desplegar cada módulo de manera independiente.
- **Integración:** permite integrar cada módulo en otros sistemas.

Sacrificios del desacoplamiento:

- **Complejidad:** se puede aumentar la complejidad del sistema.
- **Costo:** puede llegar a aumentar el costo del sistema.
- **Tiempo:** se puede aumentar el tiempo de desarrollo del sistema.

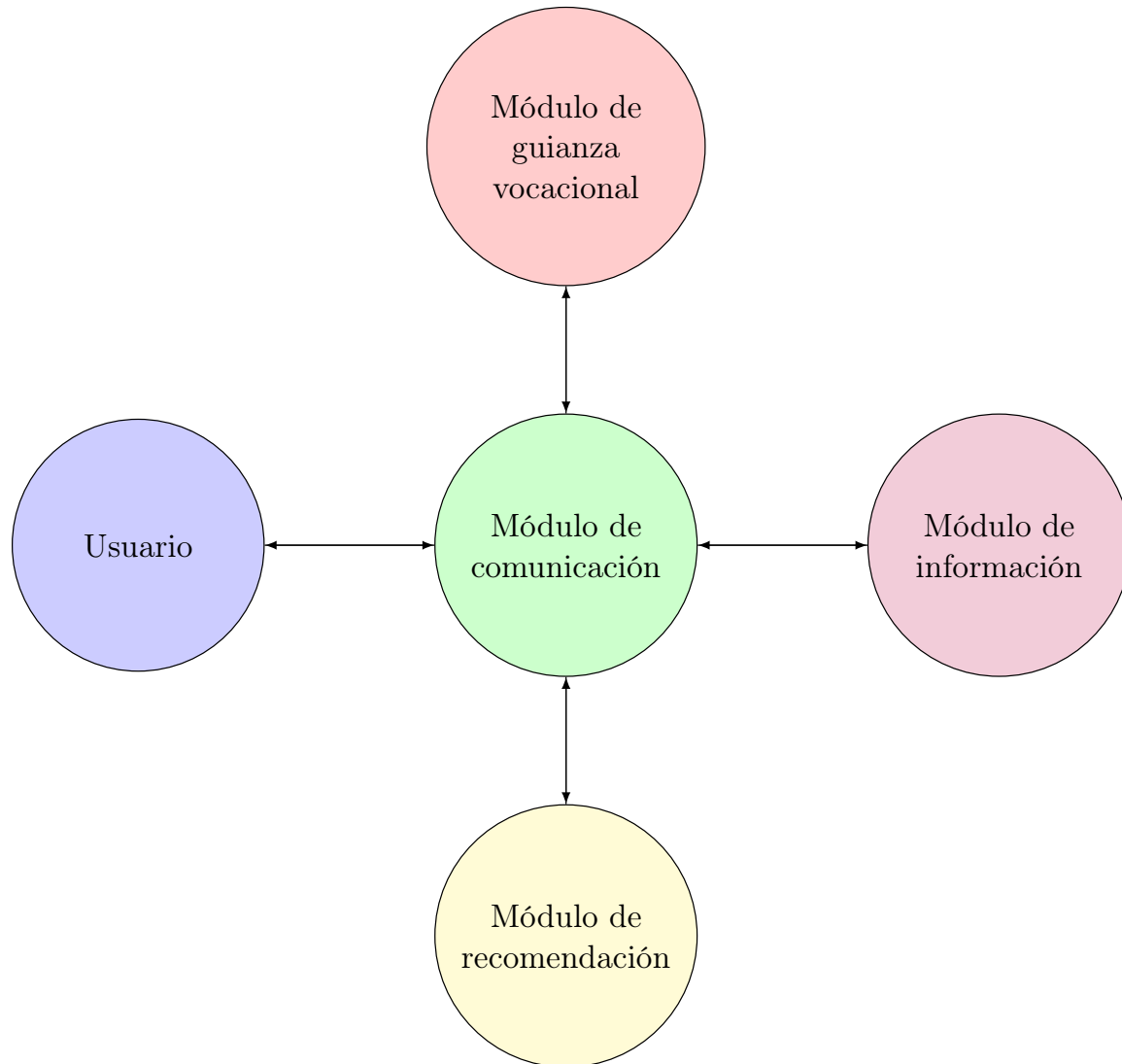
6.1.2. Arquitectura del sistema

La arquitectura propuesta plantea un sistema de 4 módulos que responden a las necesidades de proveer respuestas contextualizadas, precisas y actualizadas. Los 4 módulos principales identificados dentro del sistema son:

- **Módulo de información (sec. 6.2):** se encargará de **almacenar** la información académica sobre **carreras, cursos y pensum**, para ser utilizada por el módulo de guianza vocacional.
- **Módulo de guianza vocacional (sec. 6.3):** se encargará de **responder preguntas** relacionadas con **carreras universitarias, planes de estudio y contenidos temáticos**, utilizando información confiable y actualizada proveniente de una fuente con datos institucionales.
- **Módulo de recomendación (sec. 6.4):** se encargará de **recomendar** las **carreras universitarias y contenidos** basado en las **características de los usuarios**, utilizando los datos recolectados de una prueba de evaluación.
- **Módulo de comunicación (sec. 6.5):** se encargará de la **comunicación** entre los módulos del sistema y el **usuario**.

6.1.3. Diagrama de la arquitectura del sistema

Figura 4. Diagrama de la arquitectura del sistema propuesta



6.2. Módulo de información

El objetivo de este módulo fue construir una base de datos estructurada, consistente y semánticamente enriquecida que sirviera como fuente de conocimiento para el modelo de guianza desarrollado en etapas posteriores del proyecto. La base debía integrar información de carreras universitarias de la Universidad del Valle de Guatemala, con descripciones, plan de estudio, áreas de desarrollo, empleabilidad que facilitaran su indexación y recuperación por similitud.

6.2.1. Recopilación de datos

Fuentes de información

Las fuentes principales de información que se emplearán para la construcción de la base de datos serán páginas oficiales, institucionales y portales reconocidos por su confiabilidad y autoridad en el contexto educativo y laboral. En particular, se dará prioridad a portales universitarios oficiales, ministerios de educación, y plataformas de análisis ocupacional o salarial que cuentan con validación institucional y actualización periódica de sus contenidos.

Para la recolección sistemática de la información, se aplicaron técnicas de web scraping orientadas a la extracción automatizada de datos desde las páginas identificadas, gestionando de forma responsable el acceso y respetando archivos `robots.txt`, políticas de uso y consideraciones éticas relacionadas con el uso de datos en dominio público.

La información extraída se estructurará en formato JSON, seleccionando este formato por su flexibilidad para representar datos heterogéneos y su facilidad de integración en procesos posteriores de limpieza, enriquecimiento semántico y consulta.

6.2.2. Limpieza de datos

La limpieza de datos es un paso fundamental en cualquier proceso de gestión y análisis de información debido a varias razones críticas:

- **Eliminación de errores:** los datos recolectados, especialmente a través de técnicas automáticas como el web scraping, pueden contener errores de formato, duplicados, inconsistencias, valores faltantes o registros irrelevantes. La limpieza permite identificar y corregir estos problemas para asegurar la precisión de la base de datos.
- **Consistencia y coherencia:** es indispensable homogenizar los valores y formatos de los datos (por ejemplo, normalizar nombres de campos, unificar unidades o categorías, manejar caracteres especiales), de manera que los distintos registros sean comparables y consigan integrarse sin ambigüedades en la estructura final.
- **Facilitación de la búsqueda y recuperación:** datos limpios y consistentes mejoran los procesos de indexación y recuperación por similitud, haciendo posible consultas más efectivas y relevantes en el sistema de recomendación.
- **Mejora en el análisis y toma de decisiones:** la calidad de los análisis, modelos de recomendación y resultados finales depende directamente de la calidad de los datos de entrada. Datos limpios reducen el riesgo de sesgos, malas interpretaciones o recomendaciones poco confiables.
- **Enriquecimiento semántico:** un proceso riguroso de limpieza permite identificar y conservar la información relevante, así como enriquecer los registros con metadatos o etiquetas adicionales esenciales para el modelado semántico y funcional posterior.

6.2.3. Atributos y entidades de datos

El diseño de la base de conocimiento se fundamentó en un conjunto de requisitos funcionales y no funcionales, concebidos para asegurar la robustez, flexibilidad y escalabilidad del sistema de recomendación y guianza. La estructura de datos debe satisfacer las siguientes necesidades clave:

- **Representación integral de la información académica:** el sistema requiere un modelo capaz de almacenar no solo los datos básicos de las carreras, sino también estructuras complejas y anidadas.
- **Enriquecimiento semántico para la recomendación:** para que el motor de recomendación pueda realizar comparaciones significativas, es crucial disponer de una capa semántica que conecte las carreras con los perfiles de usuario. Se necesita un vocabulario controlado que represente habilidades, intereses y áreas de conocimiento, permitiendo un emparejamiento preciso más allá de las palabras clave.
- **Incorporación de evidencia cualitativa:** el sistema no solo debe basarse en datos estructurados, sino también en narrativas y experiencias reales. Por ello, se necesita integrar experiencias de forma estructurada, vinculándolos a sus respectivas carreras para aportar un contexto humano y cualitativo a las recomendaciones.
- **Soporte para la evaluación psicométrica:** un pilar del sistema de guianza es la evaluación del perfil del usuario. El modelo de datos debe permitir la definición declarativa y modular de pruebas psicométricas, incluyendo las reglas de puntuación, los ítems de las preguntas y los rasgos a medir, garantizando una evaluación consistente y reproducible.
- **Escalabilidad y flexibilidad del esquema:** dado que la información académica y los perfiles de usuario pueden evolucionar, se necesita un modelo de base de datos flexible y escalable. Esta elección responde a la necesidad de un esquema que pueda adaptarse a nuevos atributos y tipos de datos sin requerir migraciones complejas, asegurando la escalabilidad a futuro.
- **Comparación y recuperación de datos:** es fundamental que los datos puedan ser comparados de manera semántica, permitiendo identificar similitudes y relaciones significativas más allá de simples coincidencias textuales. Esto permite la recuperación eficiente de la información relevante para cada consulta, optimizando tanto la precisión como la rapidez en el acceso a los datos.

6.2.4. Comparativas de herramientas disponibles

Comparativa: SQL vs NoSQL

A continuación, se presenta una comparativa teórica entre bases de datos SQL y NoSQL, con el fin de determinar cuál resulta más adecuada para el sistema propuesto.

Cuadro 1. Comparativa: SQL vs NoSQL

Característica	SQL (Relacional)	NoSQL (No Relacional)
Estructura	Tablas con filas y columnas	Documentos, key-value, grafos
Esquema	Esquema rígido predefinido	Esquema flexible y dinámico
Relaciones	Establece relaciones entre datos	No establece relaciones entre datos
Consultas	Soporta consultas complejas (JOIN)	No soporta consultas complejas
Transacciones	Soporta transacciones ACID	No siempre soporta transacciones
Índices	Soporta índices	Soporta índices limitados
Escalabilidad	Escala vertical	Escala horizontal
Rendimiento	Óptimo para relaciones complejas	Óptimo para datos no estructurados
Consistencia	Garantiza consistencia inmediata	Consistencia eventual
Ejemplos	MySQL, PostgreSQL, Oracle	MongoDB, Cassandra, Redis

Nota. [128].

Comparativa: bases de datos no relacionales

A continuación, se presenta una comparativa teórica entre bases de datos No Relacionales, con el fin de determinar cuál resulta más adecuada para el sistema propuesto.

Cuadro 2. Comparativa: Bases de datos no relacionales

Característica	Documentos	Key-Value	Columna Ancha	Grafos
Ejemplo	MongoDB, DocumentDB	Redis, Memcached	Cassandra, HBase	Neo4j, ArangoDB
Estructura	JSON/BSON	Clave-valor	Filas con columnas dinámicas	Nodos y relaciones
Flexibilidad	Flexible y dinámico	No flexible	Moderadamente flexible	Flexible y dinámico
Escalabilidad	Escala horizontal	Escala horizontal	Escala horizontal	Escala horizontal
Rendimiento	Excelente lectura/escritura	Ultra rápido	Optimizado para análisis	Óptimo para relaciones
Consultas	Consultas complejas	Búsqueda simple	Agregaciones analíticas	Recorrido de grafos
Casos de Uso	Contenido, catálogos	Cache, sesiones	Análisis de datos	Redes sociales, recomendaciones
Curva de Aprendizaje	Moderada	Muy baja	Alta	Muy alta
Consistencia	Consistencia eventual	Consistencia eventual	Consistencia eventual	Fuerte consistencia

Nota. [129], [130], [131].

Comparativa: bases de datos vectoriales

A continuación, se presenta una comparativa teórica entre Bases de datos Vectoriales, con el fin de determinar cuál resulta más adecuada para el sistema propuesto.

Cuadro 3. Comparativa: Bases de datos vectoriales

Servicio	Tier Gratuito	Plan Pago	Licencia	Facilidad de Implementación
Chroma	Auto-host gratuito Cloud: \$5 créditos + 10 bases de datos	Starter: \$2.50/GiB Team: \$250/mes	Apache 2.0	Muy alta; excelente documentación; integraciones LLM listas
Pinecone	Cloud: - 2 GB almacenamiento - 2M escrituras - 1M lecturas	\$50/mes + \$0.33/GB Enterprise \$500/mes	SaaS privativo	Alta; SaaS gestionado; SDKs completos; requiere API key
Weaviate	Auto-host gratuito	Cloud ~\$25/mes Standard ~\$295/mes	Open-source	Media; GraphQL complejo; mejor para proyectos medianos
Faiss	Auto-host gratuito	N/A (librería)	MIT	Baja; requiere conocimientos avanzados; C++/Python
Qdrant	Auto-host gratuito Cloud: 1 GB	Según uso Enterprise con soporte	Rust	Media-alta; REST API simple; buena documentación

Nota. [132], [133], [134], [135], [136], [137], [138].

6.3. Módulo de guianza vocacional

El objetivo de este módulo es implementar un sistema de guianza vocacional que permita a los estudiantes explorar carreras universitarias y planes de estudio, de manera conversacional, utilizando información confiable y actualizada proveniente de las fuentes institucionales.

6.3.1. Necesidades del sistema

Para el desarrollo del sistema de guianza vocacional, se establecieron un conjunto de requisitos funcionales y no funcionales destinados a garantizar respuestas precisas, confiables y seguras. La arquitectura del sistema debe diseñarse para satisfacer las siguientes necesidades clave:

- **Precisión y verificabilidad de la información:** la necesidad primordial es que el sistema proporcione respuestas basadas principalmente en una fuente de conocimiento controlada y actualizada (la base de datos de conocimiento). Esto es crucial para evitar la generación de información incorrecta o "alucinaciones" por parte del modelo de lenguaje, garantizando que las respuestas sean siempre confiables y verificables.
- **Comprensión semántica de las consultas:** el sistema debe ser capaz de interpretar la intención del usuario más allá de las palabras clave. Se requiere un mecanismo para convertir las preguntas formuladas en lenguaje natural a un formato que permita la búsqueda por significado y contexto, no solo por coincidencia de texto.
- **Generación de respuestas coherentes y contextualizadas:** el sistema debe ser capaz de sintetizar la información recuperada y presentarla al usuario en un formato de lenguaje natural, coherente y fácil de entender. La respuesta no debe ser una simple transcripción de los datos, sino una elaboración contextualizada con el perfil del usuario.

6.3.2. Requisitos de seguridad y ética

Adicionalmente, se identificaron requisitos no funcionales críticos para asegurar un despliegue responsable y seguro:

- **Aislamiento del modelo y protección de datos:** es fundamental que el modelo de lenguaje no tenga acceso directo ni capacidad de escritura sobre la base de datos principal. Se necesita una capa de abstracción que actúe como intermediario para proteger la integridad de los datos y prevenir posibles vulnerabilidades o ataques de inyección de prompts.
- **Control del dominio de conversación:** el sistema debe garantizar que las interacciones se mantengan estrictamente dentro del ámbito académico y vocacional definido. Se requiere la implementación de mecanismos de filtrado para identificar y bloquear consultas inapropiadas o fuera de tópico, así como para revisar que las respuestas generadas sean seguras y pertinentes.

6.3.3. Comparativa de herramientas disponibles

Comparativa: RAG vs. fine-tuning

Cuadro 4. Comparativa: RAG vs. fine-tuning

Aspecto	RAG	Fine-Tuning
Actualización	Añade contexto en tiempo de inferencia desde bases externas.	Re-entrena el modelo con datos específicos del dominio.
Frescura	Información actualizable sin reentrenamiento.	Conocimiento estático; requiere reentrenamiento.
Modificación	No modifica pesos; recupera documentos relevantes.	Ajusta pesos del modelo al dominio.
Datos	Documentos sin anotación manual.	Datos curados, etiquetados y de alta calidad.
Costos	Bajo entrenamiento; requiere infraestructura de recuperación.	Alto entrenamiento; inferencia rápida después.
Ventajas	Información actual; reduce alucinaciones; trazabilidad de fuentes.	Alta precisión; control de tono/estilo; vocabulario especializado.
Limitaciones	Dependencia de infraestructura; limitada por ventana de contexto.	Riesgo sobreajuste; conocimiento estático; sin trazabilidad.

Nota. [139], [140], [141], [142].

Comparativa: modelos de embeddings

A continuación, se presenta una comparativa teórica entre Modelos de Embeddings, con el fin de determinar cuál resulta más adecuada para el sistema propuesto.

Acronyms:

RPM: Requests per Minute

RPD: Requests per Day

TPM: Tokens per Minute

CPM: Calls per Month

Cuadro 5. Comparativa: modelos de embeddings

Proveedor / Modelo	Límites de uso gratis	Costo	Rendimiento	Velocidad
OpenAI — text-embedding-3-small	- RPM: 100 - RPD: 2,000 - TPM: 40,000	US\$ 0.02 / 1M tokens	Medio (MTEB Rank 35)	Alta
OpenAI — text-embedding-3-large	- RPM: 100 - RPD: 2,000 - TPM: 40,000	US\$ 0.13 / 1M tokens	Alto (MTEB Rank 18)	Baja
Google — gemini-embedding-001	- RPM: 100 - RPD: 1,000 - TPM: 30,000	US\$ 0.15 / 1M input tokens	Alto (MTEB Rank 2)	Alta
Cohere — embed-english-v3.0	- RPM: 100 1000 CPM	US\$ 0.12 / 1M tokens	Alto (MTEB Rank 45)	Inestable (Optimizado para texto e imagenes)

Nota. [143], [144], [145], [146].

Comparativa: modelos de LLMs

A continuación, se presenta una comparativa teórica entre Modelos de LLMs, con el fin de determinar cuál resulta más adecuada para el sistema propuesto.

Cuadro 6. Comparativa: modelos de LLMs

Modelo / API	Plan gratuito	Planes pagados	Rendimiento	Velocidad
extbfGPT-3.5 Turbo (OpenAI)	Créditos iniciales (\$5-\$18); sin plan gratuito permanente.	\$0.5/1M input \$1.5/1M output	Optimizado para chat; buena relación coste-rendimiento.	Baja latencia
extbfClaude 3 Haiku (Anthropic)	Créditos iniciales (\$5); sin plan gratuito permanente.	\$0.25/1M input \$1.25/1M output	Modelo más rápido.	21K tokens/seg; 3× más veloz
extbfMistral Small v24.09 (Mistral AI)	Tier gratuito en La Platforme; prototipado sin coste.	\$0.20/1M input \$0.60/1M output	22B parámetros; mejor razonamiento.	Baja latencia
extbfCohere Command-Light	Trial API key auto-creada; rate-limited.	\$0.30/1M input \$0.60/1M output	Ligero; optimizado para velocidad/coste.	Más rápido que variantes superiores

Nota. [147], [148], [149], [150], [151].

6.4. Módulo de recomendación

El objetivo de este módulo es implementar un sistema de recomendación que pueda analizar los datos de los usuarios y recomendar tanto las carreras universitarias como el contenido temático con las características de los usuarios extraídas de una prueba de evaluación.

La metodología para el desarrollo de este módulo se divide en tres fases principales: la definición y recolección de datos para la construcción de perfiles, el desarrollo de un modelo predictivo para la recomendación de carreras, y la implementación de un algoritmo para la recomendación de contenido personalizado.

6.4.1. Definición y recolección de datos

- **Necesidad de un perfil psicométrico robusto:** para poder realizar recomendaciones significativas, el sistema requiere una comprensión profunda y multidimensional del

usuario. No es suficiente con conocer sus intereses superficiales; es necesario cuantificar sus rasgos de personalidad, habilidades cognitivas, aptitudes académicas e intereses vocacionales de una manera estandarizada y validada científicamente.

- **Selección de instrumentos:** se propone la conformación de una batería de pruebas psicométricas basada en instrumentos reconocidos internacionalmente. Esta selección se fundamenta en la necesidad de obtener un perfil de usuario integral, basado en teorías y herramientas con amplia validación empírica en el campo de la psicología y la orientación vocacional.
- **Metodología de recopilación:** se planifica un proceso de recopilación de datos con un público que nos proporcione datos confiables y completos. El procedimiento debe garantizar los más altos estándares éticos, incluyendo un consentimiento informado claro, la total anonimización de los participantes y el uso de los datos exclusivamente para fines académicos del proyecto. Se utilizará una plataforma digital para estandarizar la administración de las pruebas y la recopilación de resultados.

6.4.2. Modelo de recomendación de carreras (learning-to-rank)

- **Requisito de ranking vs. clasificación:** el objetivo no es simplemente clasificar a un estudiante en una única carrera, sino ofrecerle un listado ordenado (*ranking*) de las carreras más compatibles. Por lo tanto, se requiere un enfoque de *machine learning* especializado en problemas de ordenamiento.
- **Contingencia ante la escasez de datos:** se anticipa que la recopilación de datos podría no alcanzar el volumen necesario para entrenar un modelo robusto para las 33 carreras. Como plan de contingencia para la prueba de concepto, se contempla la **generación de datos sintéticos** para validar la viabilidad técnica del pipeline de entrenamiento y del modelo, aunque los resultados no sean generalizables.

6.4.3. Algoritmo de recomendación de contenido (filtro basado en contenido)

- **Necesidad de personalización y exploración:** una vez recomendadas las carreras, es crucial mantener al usuario comprometido y ayudarlo a explorar, presentándole contenido relevante y diverso (testimonios, detalles de cursos, etc.).
- **Requisitos adicionales del algoritmo:** el sistema debe incluir lógica para evitar la repetición de contenido ya visto y para introducir elementos de diversidad (p. ej., presentando contenido de carreras no priorizadas de forma aleatoria) para fomentar la exploración y evitar el sobreajuste de la personalización.

6.4.4. Comparativa de herramientas disponibles

Comparativa: lenguajes de programación

A continuación, se presenta una comparativa teórica entre lenguajes de programación, con el fin de determinar cuál resulta más adecuada para el entrenamiento y ejecución del modelo de machine learning.

Cuadro 7. Comparativa de lenguajes de programación para machine learning

Lenguaje	Curva de aprendizaje	Documentación	Comunidad	Velocidad
Python	Muy fácil; sintaxis legible y accesible	Muy extensa; numerosos tutoriales y librerías	Muy grande y activa	Lento de forma nativa; mitigado con librerías optimizadas y GPU
R	Moderada a empinada para no-estadísticos	Completa pero orientada a estadística	Comunidad académica fuerte, menor que Python	Ejecutado de forma interpretada; adecuado para análisis, menos para cálculos pesados
Julia	Moderada; sintaxis familiar para científicos	Documentación en crecimiento	Comunidad pequeña pero creciendo	Muy rápida gracias a su compilación JIT y eficiencia en memoria
Java	Moderada; verbose pero bien estructurado	Buena documentación oficial y en foros	Amplia en entornos empresariales	Compilado; rendimiento alto y portable; ideal para sistemas empresariales
C++	Difícil; sintaxis compleja y manejo de memoria	Documentación técnica abundante pero dispersa	Comunidad grande, centrada en sistemas	Muy rápido; se usa cuando cada microsegundo cuenta

Nota. [152], [153].

Comparativa: algoritmos de machine learning

A continuación, se presenta una comparativa teórica entre Algoritmos de Machine Learning, con el fin de determinar cuál resulta más adecuada para el entrenamiento y ejecución del modelo de machine learning.

Cuadro 8. Comparativa de varios algoritmos de *machine learning*

Algoritmo	Complejidad computacional	Requerimiento de datos	Paradigma	Casos de uso típicos
Regresión lineal	Media	Bajo–moderado	Supervisado (regresión)	Predicción de valores continuos
Regresión logística	Baja	Bajo–moderado	Supervisado (clasificación binaria)	Clasificación binaria (p.ej., detección de fraude)
SVM	Alta	Moderado–alto	Supervisado (clasificación/regresión)	Clasificación compleja, separación de clases
Árboles de decisión	Media	Moderado	Supervisado	Modelos interpretables, segmentación
Bosques aleatorios	Media–alta	Moderado–alto	Supervisado	Clasificación robusta, ranking de características
k-NN	Media–alta	Moderado–alto	Supervisado	Clasificación basada en similitud
Naive Bayes	Baja	Bajo	Supervisado	Clasificación de texto, filtrado de spam
k-means	Media	Moderado–alto	No supervisado (clustering)	Segmentación, agrupamiento exploratorio
Redes neuronales	Alta	Alto	Supervisado (o semi-supervisado)	Reconocimiento de voz/imágenes, modelado de lenguaje

Nota. [154], [155], [156].

Comparativa: algoritmos de árboles de decisión

Cuadro 9. Comparativa de algoritmos de árboles de decisión

Algoritmo	Casos de uso típicos (generales)	Coste de ejecución	Otras características relevantes
Árbol de decisión (simple)	Clasificación y regresión sencillas; ideal como modelo base o para ranking básico	Bajo. Entrenamiento y predicción rápidos; pocos recursos computacionales	Alta interpretabilidad; maneja datos numéricos y categóricos; robusto a outliers y fácil de actualizar; susceptible a sobreajuste si no se poda
Bosque aleatorio	Clasificación, regresión y ranking con conjuntos de datos grandes o complejos	Moderado. Entrena múltiples árboles en paralelo; escalable y eficiente en datasets grandes	Menos propenso al sobreajuste; proporciona importancia de variables; requiere más memoria; predicciones más lentas por el voto de varios árboles
Gradient Boosting (GBT / XGBoost)	Clasificación y regresión de alta precisión; ranking y tareas con datos desbalanceados	Alto. Entrena árboles de forma secuencial corrigiendo errores previos, con tiempos de entrenamiento largos	Muy preciso y flexible; sensible a hiperparámetros; mayor riesgo de sobreajuste; menos interpretable
Extra Trees (Extremely Randomized Trees)	Clasificación y regresión rápidas; problemas de ranking y clasificación en alta dimensión con poco coste computacional	Bajo-moderado. Usa el conjunto completo de datos y selecciona puntos de corte aleatorios, reduciendo el tiempo de entrenamiento	Mayor aleatorización reduce varianza y sesgo; similar rendimiento al Random Forest pero más rápido; adecuado cuando la velocidad es prioritaria

Nota. [164], [165], [166].

6.5. Módulo de comunicación

La API de comunicación es la interfaz entre el sistema y el usuario, y entre los diferentes módulos del sistema. Para ello, se propone la implementación de una API RESTful que permita conectar al analizar y procesar las diferentes interacciones entre el sistema y el usuario.

6.5.1. Arquitectura y principios de diseño

Para cumplir su rol como núcleo de la interacción, el diseño del Módulo de Comunicación se basa en tres principios fundamentales: orquestación centralizada (sec. 5.6.1), arquitectura orientada a servicios (sec. 5.6.2) y una selección tecnológica que garantice robustez y rendimiento.

- **Orquestación centralizada:** la API no actúa como un simple intermediario, sino como un **orquestador** activo. Es responsable de gestionar los flujos de trabajo complejos que involucran múltiples componentes, como la consulta a bases de datos documentales y vectoriales, y la comunicación con los modelos de lenguaje (LLM). Esto centraliza la lógica de negocio y simplifica la interacción desde el cliente.
- **Arquitectura orientada a servicios:** se adopta un enfoque de servicios donde la API expone un conjunto de *endpoints* cohesivos y especializados. Cada grupo de *endpoints* corresponde a una capacidad de negocio específica, encapsulando la lógica de los módulos subyacentes. Este diseño promueve la modularidad y facilita el mantenimiento y la escalabilidad del sistema.

6.5.2. Necesidades funcionales y flujos de datos

Basado en estos principios, el módulo debe satisfacer las siguientes necesidades funcionales, cada una asociada a un flujo de orquestación específico:

- **Gestión de la evaluación inicial:** orquestar la obtención de los ítems de la prueba psicométrica y procesar las respuestas del usuario para invocar al módulo de recomendación y devolver un perfil inicial de carreras.
- **Guianza conversacional:** facilitar la interacción conversacional al: entender la información necesaria partiendo de la pregunta del usuario, recuperar la información relevante de la base de datos de información académica y generar una respuesta informada.
- **Generación de contenido temático:** orquestar la creación de contenido personalizado, mini-informes, tarjetas de información y preguntas de la generación de contenido temático, gestionando las llamadas a los servicios de generación de contenido correspondientes con información de la base de datos de información académica.

- **Recomendación de carreras similares:** exponer una interfaz que pueda encontrar carreras similares a la carrera buscada por el usuario, utilizando lógica de similitud de los datos de la base de datos de información académica.

De esta forma, el módulo de comunicación no solo debe exponer funcionalidades, sino que debe diseñar una arquitectura robusta que garantice la correcta colaboración entre los distintos componentes especializados del sistema.

6.5.3. Comparativa de herramientas disponibles

Comparativa: API RESTful vs. API GraphQL

A continuación, se presenta una comparativa teórica entre API RESTful y API GraphQL, con el fin de determinar cuál resulta más adecuada para el sistema propuesto.

Cuadro 10. Comparación de las arquitecturas REST y GraphQL

Característica	REST	GraphQL
Modelo de datos y puntos de acceso	Múltiples endpoints basados en recursos; devuelve estructuras fijas	Un único endpoint; cliente especifica los campos que desea
Caché	Mecanismos HTTP nativos (ETags, Cache-Control)	Caching más complejo; requiere estrategias personalizadas
Versionado	API versionada mediante rutas (/v1/, /v2/)	Evolución del esquema sin versiones explícitas
Eficiencia y complejidad	Escalable; bueno para CRUD simples	Reduce llamadas y carga (30–50 %); puede ser cuello de botella en consultas complejas
Casos de uso típicos	APIs públicas, microservicios, catálogos de productos	Aplicaciones con datos complejos, dashboards, móvil con ancho limitado
Seguridad y complejidad operativa	Autenticación / autorización bien establecidas	Control de acceso complejo; estimación de costes difícil
Curva de Aprendizaje	Baja	Media
Documentación	Buena	Buena
Rendimiento	Media	Alta
Costo	Bajo	Alto

Nota. [161], [162], [163].

Comparativa: lenguajes de programación para la implementación de la API

A continuación, se presenta una comparativa teórica entre lenguajes de programación para la implementación de la API, con el fin de determinar cuál resulta más adecuada para el sistema propuesto.

Cuadro 11. Comparativa de lenguajes de programación para la implementación de la API

Lenguaje	Simplicidad y curva de aprendizaje	Ecosistema y frameworks	Rendimiento y escalabilidad	Casos de uso típicos
JavaScript/Node.js	Mismo lenguaje frontend/backend; sintaxis sencilla	Ecosistema enorme (NPM); frameworks: Express, NestJS, Fastify	Maneja muchas peticiones con bajo coste; no ideal para tareas CPU-intensivas	APIs web/móviles, microservicios ligeros
TypeScript	Tipos estáticos sobre JavaScript; mejora mantenibilidad	Ecosistema Node; frameworks: NestJS, TypeORM	Compila a JavaScript sin coste de rendimiento	APIs empresariales, proyectos grandes
Python	Sintaxis legible; ideal para principiantes	FastAPI, Flask, Django REST; excelente para datos y ML	Interpretado; suficiente para APIs; fácil escalado horizontal	APIs de datos, inferencia ML, automatización
Java	Sintaxis verbosa; comunidad enorme y abundantes recursos	Spring Boot, Jakarta EE	Compilado; alto rendimiento, multithreading, seguridad	APIs empresariales, sistemas B2B de alto tráfico
Go (Golang)	Sintaxis minimalista; fácil de aprender	Gin, Echo, Fiber; binarios compilados; integración gRPC	Muy rápido; concurrencia incorporada con goroutines	Microservicios, sistemas distribuidos, Kubernetes
C# (.NET)	Similar a Java; más rápido con entorno Microsoft	ASP.NET Core; Visual Studio, Azure DevOps	Compilado; buen rendimiento; ideal para Windows/Azure	CRM/ERP empresariales, Azure, backends de videojuegos
PHP	Curva de aprendizaje corta; resultados rápidos	Laravel, Symfony; comunidad amplia	Interpretado; tráfico moderado	CMS, e-commerce, aplicaciones web tradicionales
Ruby	Sintaxis limpia; gran productividad	Ruby on Rails; desarrollo rápido	Interpretado; escalabilidad menor	Prototipado rápido, startups, MVPs
Rust	Alta complejidad inicial; memoria segura	Actix Web, Rocket	Compilado; alto rendimiento, baja latencia	Servicios críticos en seguridad y eficiencia

Nota. [164], [165], [166].

Comparativa: frameworks TypeScript para la implementación de la API

Cuadro 12. Comparativa de frameworks TypeScript para APIs RESTful

Framework	Casos de uso (tamaño del proyecto)	Rendimiento	Curva de aprendizaje	Poder de cómputo necesario
Express.js	Proyectos pequeños y medianos, prototipos	Moderado	Muy baja	Bajo
Koa.js	APIs ligeras, control total	Alto	Moderada	Bajo
Fastify	APIs alto rendimiento, tiempo real	Muy alto	Moderada	Moderado
NestJS	Aplicaciones empresariales de gran escala	Moderado	Alta	Alto
Hono	Proyectos pequeños, edge/serverless	Alto	Baja	Muy bajo
Elysia	APIs con tipado estricto	Alto	Moderada	Moderado

Nota. [167], [168], [169].

7.1. Resultados encuesta inicial

Se realizó una encuesta inicial dirigida a estudiantes de nivel medio (bachillerato), con el propósito de recopilar información sobre sus intereses, percepciones y necesidades en torno a la elección de una carrera universitaria. El cuestionario incluyó preguntas demográficas básicas, el nivel de conocimiento que poseen acerca de sus opciones académicas, las fuentes de información a las que recurren, el grado de orientación recibida en su centro de estudios, así como las principales barreras y emociones asociadas al proceso de decisión vocacional.

Además, se indagó en las expectativas de los estudiantes respecto a una plataforma digital de orientación vocacional, consultando qué características consideran más relevantes (como recibir recomendaciones personalizadas, comparar planes de estudio, guardar opciones de interés o interactuar con un asistente virtual).

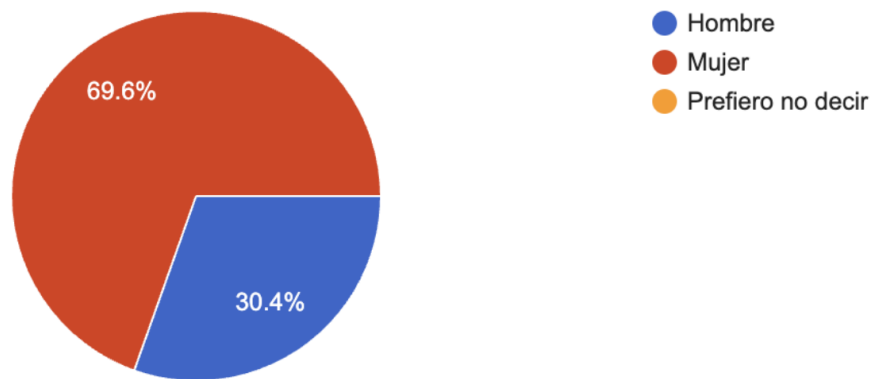
Los resultados de esta encuesta constituyen una línea base para la investigación, ya que permiten identificar los factores más influyentes en la elección de carrera y sirven como insumo para el diseño del modelo de recomendación y el sistema de guianza vocacional basado en inteligencia artificial.

7.1.1. Perfil de usuarios encuestados

Distribución de género

La distribución de género de los usuarios encuestados, nos muestra que la mayoría de los usuarios encuestados son hombres. Por lo que nuestros datos pueden estar sesgados hacia las preferencias y expectativas del género masculino.

Figura 5. Distribución de género de los usuarios encuestados



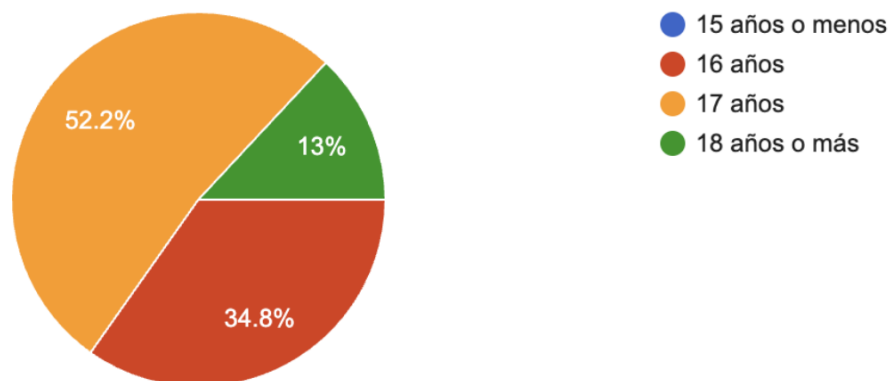
Nota. Elaboración propia.

Distribución de edades

La distribución de edades de los usuarios encuestados, nos muestra que la mayoría de los usuarios encuestados se encuentran entre las edades de 15 y 18 años. Por la proporción de las edades, se puede inferir que la mayoría de encuestados se encuentran dentro de las últimas etapas de bachillerato, siendo la mayor población de la edad de 17 años.

Esto es importante ya que nuestro proyecto se enfoca en estudiantes de bachillerato, por lo que es importante que el modelo de recomendación y el sistema de guianza vocacional basado en inteligencia artificial sean diseñados para este grupo de edad.

Figura 6. Distribución de edades de los usuarios encuestados

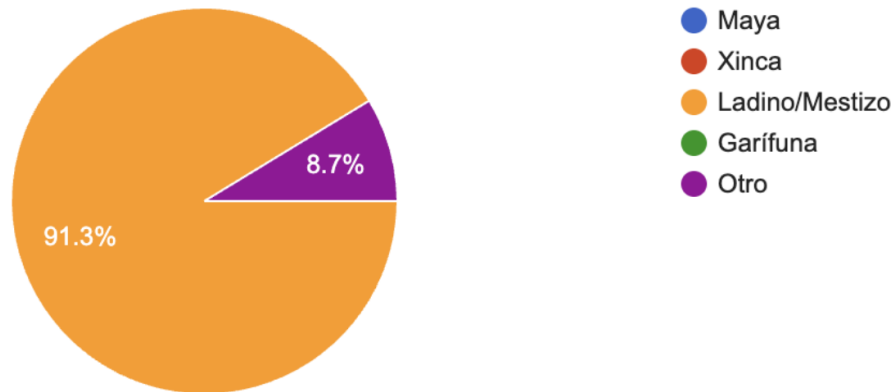


Nota. Elaboración propia.

Distribución de etnia

La distribución observada puede explicarse en función de la composición demográfica de la población encuestada. En este caso, la muestra está conformada principalmente por estudiantes de la Ciudad de Guatemala, donde la población tiende a ser relativamente homogénea en términos étnicos, predominando la etnia ladina/mestiza.

Figura 7. Distribución de etnicidad de los usuarios encuestados

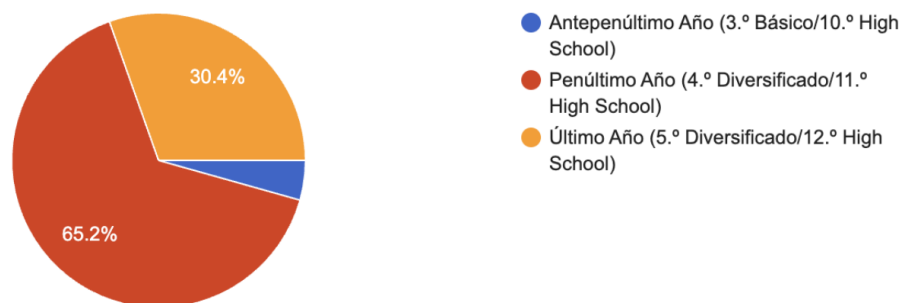


Nota. Elaboración propia.

Distribución de grados escolares

La distribución de grados escolares de los usuarios encuestados, nos muestra que la mayoría de los usuarios encuestados son de bachillerato, incluyendo algunos de últimos años de básicos. Podemos ver que la gran mayoría de encuestados, con un (65.2 %), se encuentran cursando la última etapa de bachillerato.

Figura 8. Distribución de grados de los usuarios encuestados

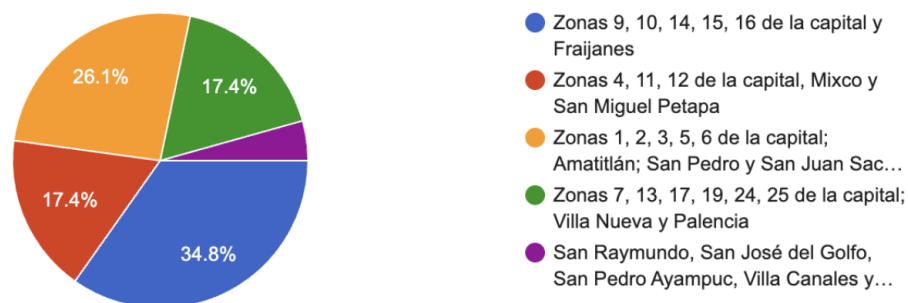


Nota. Elaboración propia.

Distribución de localización de centros de estudios

La distribución de localización de centros de estudios de los usuarios encuestados, nos muestra que una gran parte de los encuestados estudian en instituciones dentro de la Ciudad Capital de Guatemala, incluyendo ciudades y comunidades aledañas.

Figura 9. Distribución de localización de centros de estudios de los usuarios encuestados

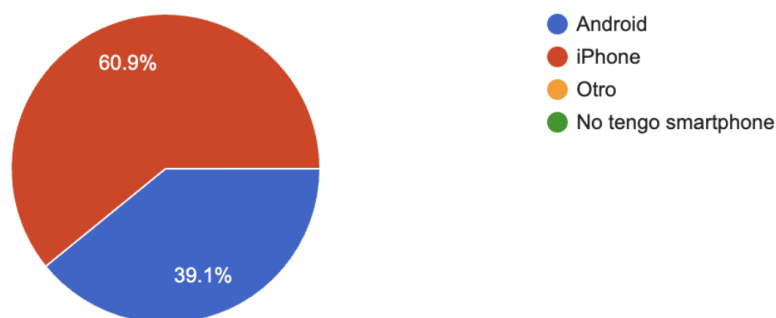


Nota. Elaboración propia.

Distribución de disponibilidad de un dispositivo digital

La gráfica muestra que todos los encuestados poseen un dispositivo digital, siendo esto una gran ventaja para el proyecto, ya que, gracias a los avances tecnológicos, esto permite que cualquier estudiante pueda acceder a cualquier plataforma o aplicación.

Figura 10. Distribución de marca de teléfono de los usuarios encuestados



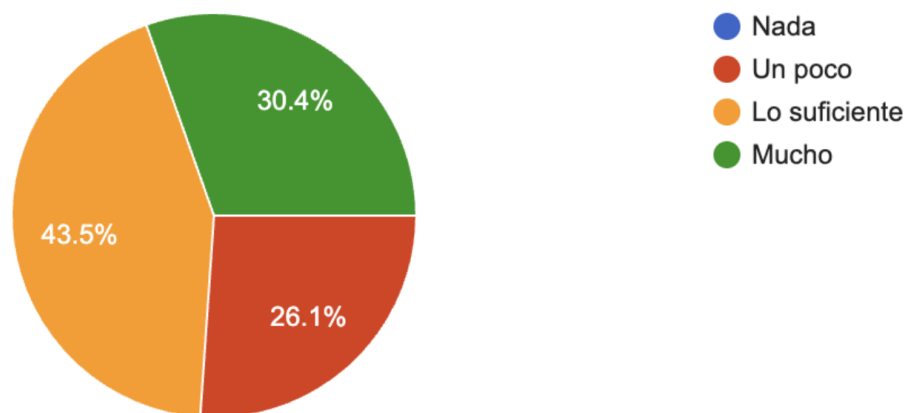
Nota. Elaboración propia.

7.1.2. Exploración y apoyo en futuro profesional

Distribución primera pregunta: ¿Entiendes tus opciones de carrera?

Al observar la distribución, se puede observar que la gran mayoría de encuestados posee una buena idea de sus opciones de carrera, juntando un (56.5 %); sin embargo, se observa que un (43.5 %) no posee una clara idea de sus opciones de carrera.

Figura 11. Distribución primera pregunta de apoyo futuro de los usuarios encuestados

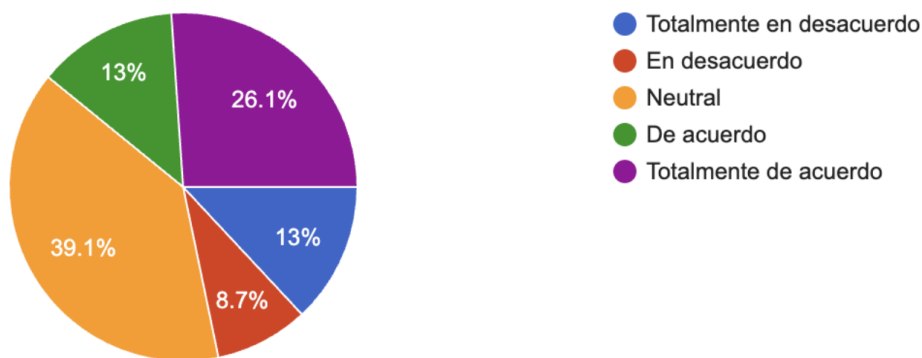


Nota. Elaboración propia.

Distribución segunda pregunta: ¿Estás de acuerdo con la siguiente afirmación?: "Mi colegio me ha proporcionado suficientes herramientas y orientación para tomar una decisión informada sobre mi futuro profesional"

La distribución nos muestra que solo el (26 %) de los encuestados consideran que su centro educativo les ha proporcionado la suficiente exposición y orientación para tomar una decisión informada sobre su futuro profesional. Mientras que un (74 %) no consideran que su centro educativo ha hecho un buen trabajo en esta área.

Figura 12. Distribución segunda pregunta de apoyo futuro de los usuarios encuestados

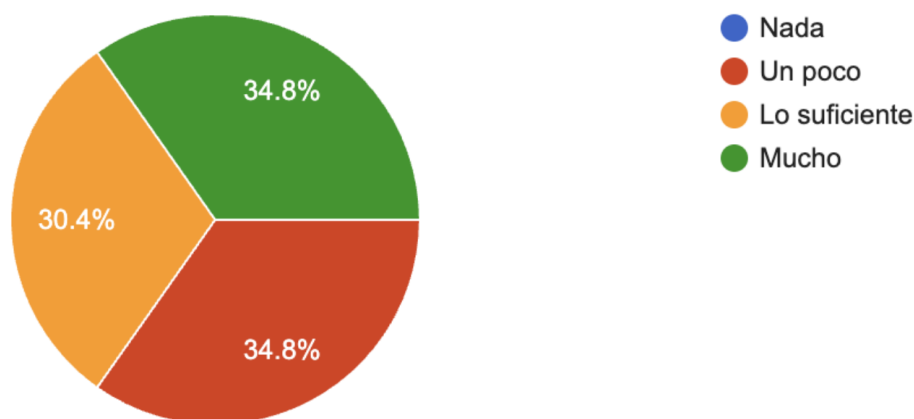


Nota. Elaboración propia.

Distribución tercera pregunta: ¿Cuánto estrés te genera pensar en la elección de tu carrera profesional?

El gráfico muestra el nivel el nivel de estrés que genera pensar en la elección de su carrera profesional a los encuestados. Los resultados muestran que un (65.2 %) de los encuestados posee una cantidad considerable de estrés por la decisión, mientras que el resto posee una cantidad controlable. Cabe recalcar que ningún estudiante declaró no tener estrés alguno por la decisión, lo cuál es entendible, considerando el grado de impacto que dicha decisión tiene en su vida.

Figura 13. Distribución tercera pregunta de apoyo futuro de los usuarios encuestados



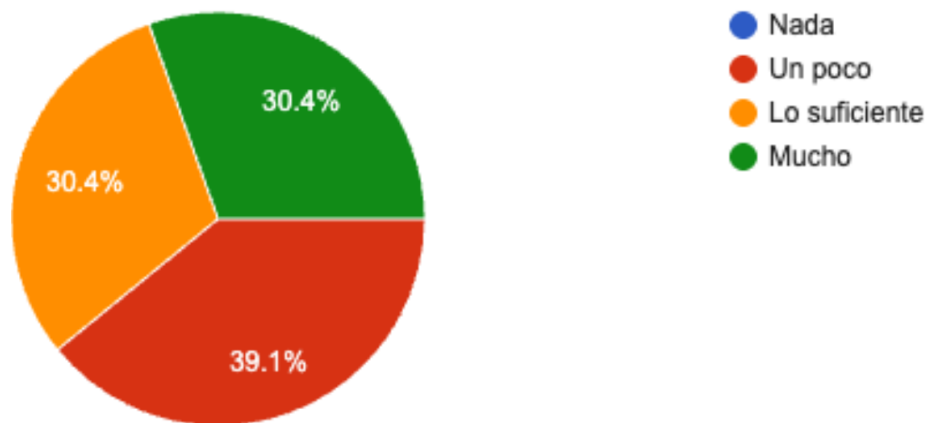
Nota. Elaboración propia.

7.1.3. Búsqueda de información y recursos

Distribución primera pregunta: ¿Has investigado activamente sobre carreras universitarias?

La distribución de la gráfica muestra que la gran mayoría de los encuestados (60.8 %) han investigado activamente sobre carreras universitarias hasta cierto grado aceptable, mientras que el resto (39.2 %) no han investigado y explorado de manera profunda sobre carreras universitarias.

Figura 14. Distribución primera pregunta de búsqueda de recursos de los usuarios encuestados

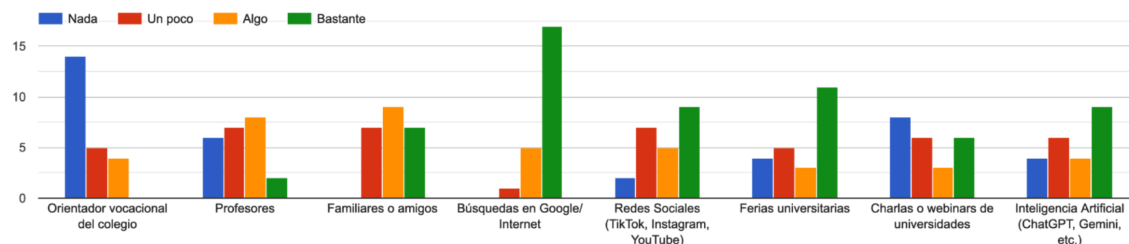


Nota. Elaboración propia.

Distribución segunda pregunta: Indica con qué frecuencia has utilizado las siguientes fuentes para informarte sobre carreras universitarias en el último año (consciente o inconscientemente)

Los resultados muestran que las fuentes más consultadas por los estudiantes para informarse sobre carreras universitarias son las búsquedas en Google/Internet y las ferias universitarias, seguidas de las redes sociales y el apoyo de familiares o amigos. Por el contrario, se puede ver que métodos más tradicionales como los orientadores vocacionales y los profesores se reportan como las fuentes menos utilizadas.

Figura 15. Distribución segunda pregunta de búsqueda de recursos de los usuarios encuestados

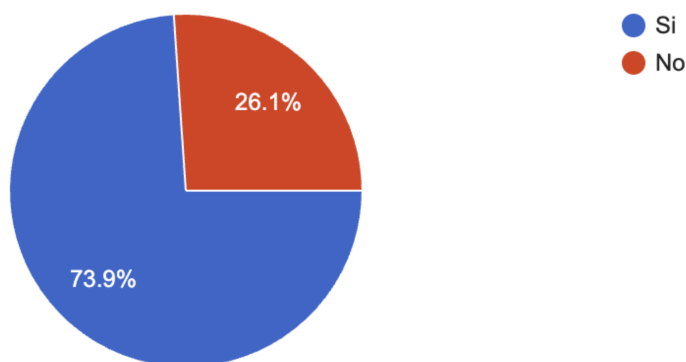


Nota. Elaboración propia.

Distribución tercera pregunta: ¿Sientes que te falta información o guía en algún aspecto?

La distribución de la gráfica muestra que la gran mayoría de los encuestados (73.9 %) sienten que les falta información o guía en algún aspecto, mientras que el resto (26.1 %) no sienten que les falta información o guía en algún aspecto.

Figura 16. Distribución tercera pregunta de búsqueda de recursos de los usuarios encuestados



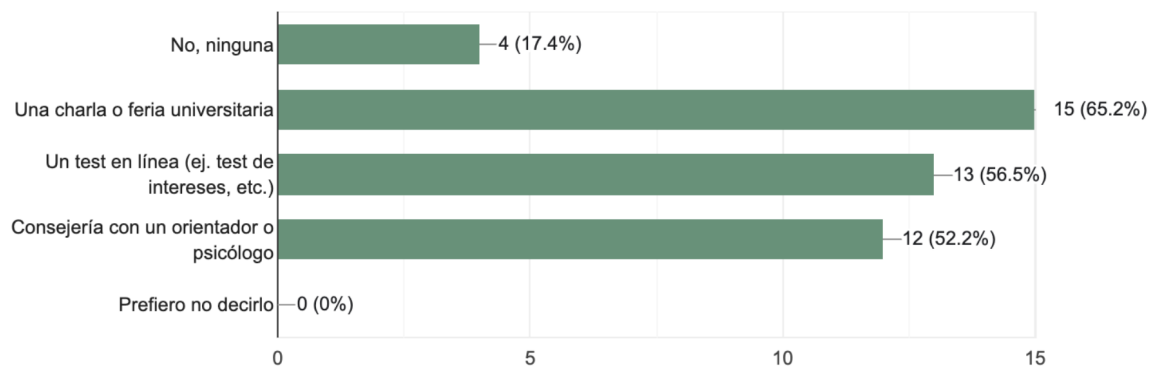
Nota. Elaboración propia.

7.1.4. Expectativas sobre herramientas de orientación vocacional

Distribución primera pregunta: Antes de hoy, ¿has usado alguna de estas herramientas de orientación vocacional?

Al preguntar a los estudiantes sobre el uso previo de herramientas de orientación vocacional, el (65.2 %) indicó haber participado en charlas o ferias universitarias, el (56.5 %) afirmó haber realizado algún test en línea y el (52.2 %) reportó haber recibido consejería con un orientador o psicólogo. Sin embargo, un (17.4 %) manifestó no haber tenido acceso a ninguna herramienta de este tipo.

Figura 17. Distribución primera pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados

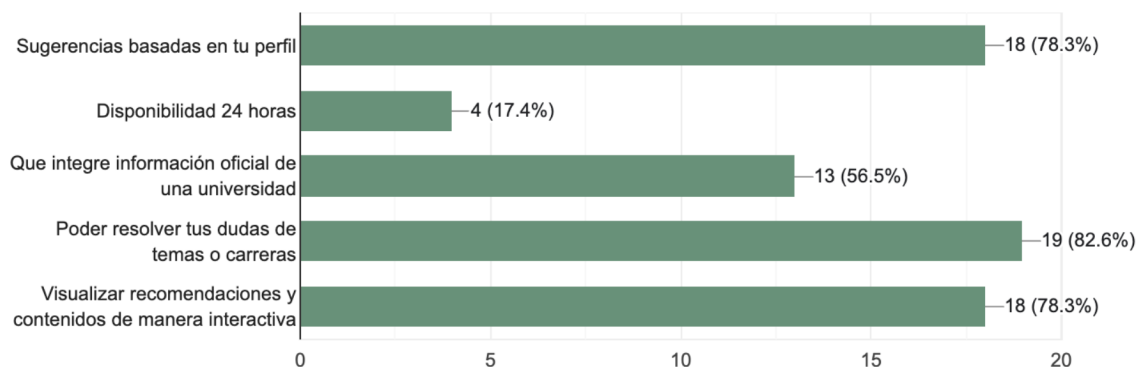


Nota. Elaboración propia.

Distribución segunda pregunta: ¿Qué características crees imprescindibles en una herramienta de orientación vocacional?

Los estudiantes consideran que las características imprescindibles de una herramienta de orientación vocacional son la posibilidad de resolver dudas sobre carreras (82.6 %), recibir sugerencias basadas en su perfil (78.3 %) y acceder a recomendaciones y contenidos de manera interactiva (78.3 %). Asimismo, un (56.5 %) destacó la relevancia de contar con información oficial de universidades, mientras que solo un (17.4 %) consideró esencial la disponibilidad las 24 horas.

Figura 18. Distribución segunda pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados

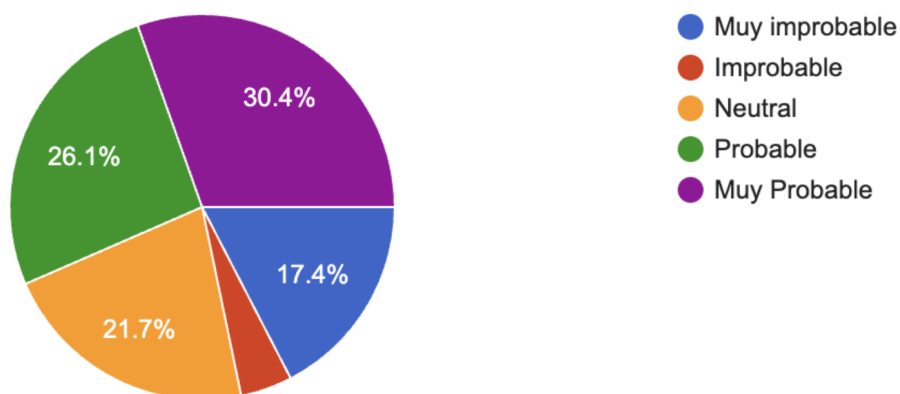


Nota. Elaboración propia.

Distribución tercera pregunta: ¿Qué tan probable sería que utilizaras una plataforma que ayude a explorar diferentes campos profesionales?

Al consultar la probabilidad de utilizar una plataforma que permita explorar diferentes campos profesionales, el (56.5 %) de los estudiantes manifestó una disposición positiva (Probable o Muy probable), mientras que un (21.7 %) se mostró neutral y únicamente un (21.7 %) expresó una postura negativa.

Figura 19. Distribución tercera pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados



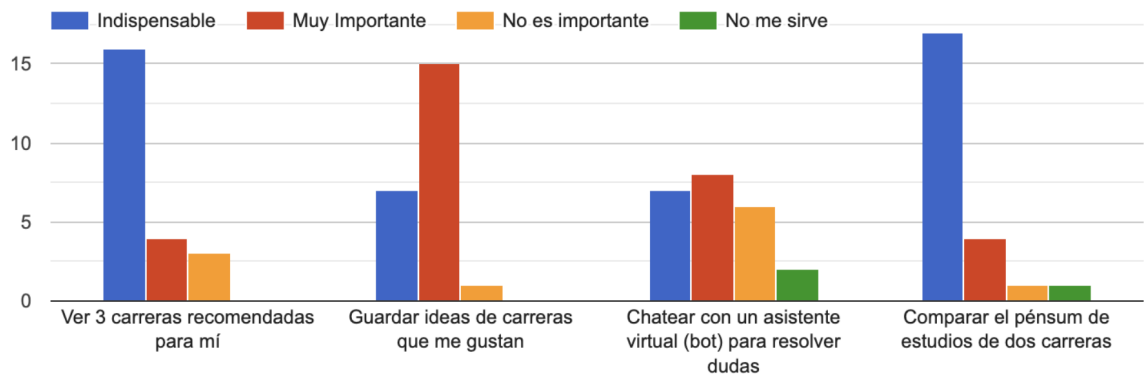
Nota. Elaboración propia.

Distribución cuarta pregunta: Si la app te ofreciera estas opciones ahora mismo, ¿cual te parecería más útil?

Al consultar qué opciones dentro de una aplicación de orientación vocacional serían más útiles, los estudiantes priorizaron dos funcionalidades: ver 3 carreras recomendadas personalizadas y comparar el pènsum de estudios de dos carreras, ambas consideradas indispensables por la mayoría. Asimismo, la opción de guardar ideas de carreras de interés fue evaluada como muy importante por más de la mitad de los participantes.

En contraste, la función de chatear con un asistente virtual obtuvo valoraciones más divididas: mientras algunos la consideran indispensable, otros manifestaron que no les resulta relevante.

Figura 20. Distribución cuarta pregunta de expectativas sobre herramientas de orientación vocacional de los usuarios encuestados

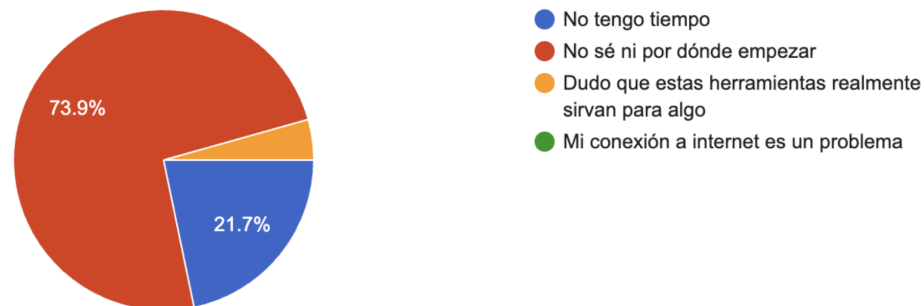


Nota. Elaboraci n propia.

Distribuci n quinta pregunta:  Cu l es la barrera m s grande que tienes ahora para investigar sobre tu futuro profesional?

Los resultados muestran que la principal barrera que enfrentan los estudiantes al investigar sobre su futuro profesional es “no saber por d nde empezar” (73.9%), seguida de “no tener tiempo” (21.7%). Un (4.3%) expres  dudas sobre la utilidad real de las herramientas existentes, mientras que ninguno mencion  problemas de conectividad como limitante.

Figura 21. Distribuci n quinta pregunta de expectativas sobre herramientas de orientaci n vocacional de los usuarios encuestados



Nota. Elaboraci n propia.

7.2. Módulo de información (obj. específicos 1 y 2 - 3.2)

El módulo de información es el encargado de la recopilación, limpieza y almacenamiento de la información de las carreras profesionales de la Universidad del Valle de Guatemala. Para ello, se propone la implementación de una base de datos no relacional utilizando una base de datos **no relacional** (sec. 5.17.7) (AWS Document DB sec. 5.17.7) para catalogar y estructurar la información de las carreras profesionales, y una base de **datos vectorial** (sec. 5.17.6) (Pinecone) (sec. 5.17.6) para la recuperación semántica de información mediante búsquedas por similitud.

Para más información sobre la estructura de la información recopilada, se puede consultar el anexo 11.1.1.

7.2.1. Justificación de elecciones técnicas

Justificación de los parámetros vectoriales

La **dimensionalidad de 768** fue seleccionada para mantener un equilibrio óptimo entre la capacidad de representación semántica y la eficiencia computacional. Esta dimensión es el estándar utilizado por modelos de embeddings ampliamente reconocidos en la industria, como BERT y sus variantes multidiollos, que han demostrado ser efectivos para tareas de similitud semántica en lenguaje natural. Una dimensión menor comprometería la precisión de las representaciones, mientras que una mayor aumentaría significativamente los costos computacionales sin aportes substanciales de precisión.

Se optó por **vectores densos** en lugar de vectores dispersos (sparse vectors) porque proporcionan una representación más compacta y eficiente de la información semántica. Los vectores densos permiten capturar relaciones semánticas profundas entre las carreras y las consultas de los usuarios, lo cual es crítico para un sistema de recomendación personalizado. Además, Pinecone optimiza el almacenamiento y la búsqueda de vectores densos, resultando en tiempos de latencia más bajos.

La **función cosine similarity** fue elegida como métrica de similitud porque es invariante a la magnitud de los vectores, permitiendo comparar direcciones de vectores normalizados. Esta propiedad es ideal para la búsqueda semántica, ya que el significado de un documento está más representado por su dirección en el espacio vectorial que por su magnitud. La cosine similarity es estándar en aplicaciones de búsqueda de similitud semántica y es especialmente efectiva para identificar carreras y testimonios que mejor se alinean con los intereses y características del usuario.

Justificación de la elección de NoSQL

Con base en la comparativa de NoSQL (sec. 1), se seleccionó nosql como tecnología de base de datos, debido a las siguientes razones:

- **Flexibilidad** en la estructura de los datos.

- No requiere de transacciones **ACID**.
- **Minimiza tiempo de desarrollo**, al no tener que diseñar esquemas rígidos.
- NoSQL **sacrifican ACID por rendimiento**, por lo que se puede obtener una respuesta más rápida a las consultas.
- Operaciones de transacciones **ACID no son necesarias**, ya que la información almacenada no requiere de actualizaciones frecuentes.
- **NoSQL posee mejor rendimiento en operaciones de lectura**, ya que no requiere de JOINS para obtener la información; agregando que nuestra información con escasa frecuencia requiere de conexiones con otras colecciones.
- **Mejor manejo de grandes cantidades a largo plazo** gracias a su escalabilidad horizontal permite aumentar la capacidad de la base de datos sin sacrificar el rendimiento.

En consecuencia, se seleccionó NoSQL como tecnología de base de datos, debido a su flexibilidad estructural, escalabilidad horizontal y rendimiento superior para la naturaleza no relacional de los datos del proyecto.

Justificación de la elección de documentos

Con base en la comparativa de Documentos (sec. 2), se seleccionó documentos como tecnología de base de datos, debido a las siguientes razones:

- Utilizan **JSON** como formato de almacenamiento para estructurar la información.
- **Documentos** son ideales para almacenar contenido libre, como descripciones de carreras, testimonios de egresados, etc.
- Al ser nuestra información contenido académico, entra dentro del caso de uso indicado.
- No dependen de relaciones complejas, ni esquemas estrictos para su almacenamiento.
- Su **curva de aprendizaje** es moderada, sin embargo, su lenguaje de consulta, similar a Javascript, hace que un cualquier persona con la suficiente experiencia en el lenguaje pueda utilizarlo sin dificultades.
- Al utilizar documento, se puede usar la redundancia de información para disminuir el tiempo de respuesta para datos con una alta frecuencia de consulta.

En consecuencia, se seleccionó Documentos como tecnología de base de datos, debido a su flexibilidad estructural, escalabilidad horizontal y rendimiento superior para la naturaleza no relacional de los datos del proyecto.

Justificación de la elección de Pinecone

Con base en la comparativa de Pinecone (sec. 3), se seleccionó Pinecone como tecnología de base de datos vectorial, debido a las siguientes razones:

- **Facilidad:** excelente documentación y SDKs en múltiples lenguajes para integración rápida.
- **Costo:** servicio gestionado; solo paga por uso sin gestionar infraestructura.
- **Plan gratuito:** 2 GB almacenamiento, 2M escrituras, 1M lecturas; suficiente para el proyecto.
- **Interfaz:** panel web intuitivo para gestionar bases de datos vectoriales.

En consecuencia, se seleccionó Pinecone por su facilidad de implementación, límites generosos en el plan gratuito e interfaz web amigable.

7.2.2. Desarrollo

Recopilación de datos

La información se consultó de fuentes oficiales y confiables pertenecientes a la **Universidad del Valle de Guatemala** [170], para la información oficial de las carreras universitarias y sitios de referencia para salarios como **TuSalario.com** [171] para Guatemala y **Salary.com** [172] para búsqueda de salarios en Estados Unidos para carreras que no tengan información en Guatemala. Para la recolección y extracción se utilizaron técnicas de web scraping y revisión manual. Estas fuentes proporcionaron datos sobre las características académicas y profesionales de distintas carreras universitarias, incluyendo:

- Nombre y plan de estudio.
- Facultad o unidad académica correspondiente.
- Descripción de contenidos, duración y perfil del egresado.
- Empleabilidad y Rango de Salarios

Limpieza de datos

Una vez recopilada la información, se realizó la **limpieza de los datos (sec. 5.4.1)**, donde se identificaron algunos errores ortográficos, dirigidos principalmente hacia la falta de tildes en algunas palabras y algunos caracteres ajenos al lenguaje español. Por lo que se aplicó un proceso de preprocesamiento lingüístico orientado a eliminar las discrepancias entre palabras y garantizar la coherencia entre los registros. Las operaciones realizadas fueron:

1. Eliminación de caracteres especiales y etiquetas HTML.
2. Conversión a minúsculas y normalización de acentos (Principalmente en títulos y textos de identificación).
3. Validación de campos obligatorios y eliminación de duplicados.

El preprocesamiento de texto es una etapa fundamental en sistemas basados en embeddings, ya que influye directamente en la calidad semántica de las representaciones vectoriales.

7.2.3. Atributos y entidades de datos documentales

La base de conocimiento se almacenó dentro de una base de datos **No SQL** (sec. 1) basada en **documentos** hospedada en **DocumentDB** (sec. 2). La información se estructuró en colecciones documentales, de forma complementaria, el contenido académico, las evidencias narrativas y la configuración psicométrica necesaria para el sistema de recomendación. A continuación se presenta una visión general de las colecciones de la base de conocimiento y su propósito dentro del flujo del sistema:

- **Careers (carreras)**: núcleo del conocimiento. Contiene la descripción de cada carrera (facultad, duración, perfil del egresado, áreas de desarrollo), el plan de estudio anidado y un arreglo de *tags* con puntajes de relevancia. Esta estructura habilita consultas semánticas y filtrados por atributos académicos.
- **Courses (cursos)**: catálogo de materias con prerequisites referenciados por `ObjectId`. Complementa a *Careers* al permitir reconstruir rutas curriculares y dependencias académicas.
- **Testimonies (testimonios)**: evidencias cualitativas de egresados, con metadatos básicos. Aportan contexto narrativo útil para enriquecer la recuperación de información y la explicación de recomendaciones.
- **Cards (tarjetas)**: capa de presentación unificada. Cada tarjeta declara su `type`, el `display_data` y una `reference` a la colección y documento de origen. Facilita componer interfaces y resultados del sistema (p. ej., mostrar una carrera o un testimonio) sin acoplarse al esquema interno.
- **Scoring rules (reglas de puntuación)**: definen, por sección de prueba (p. ej., RIA-SEC), los rasgos a medir, sus preguntas asociadas y los ítems invertidos (*reverse_scored*). Son la base para transformar respuestas en puntajes psicométricos consistentes.
- **Tags (etiquetas)**: vocabulario controlado de conceptos, habilidades e intereses. Operan como pivote semántico que enlaza carreras y testimonios con el perfil del usuario; son críticos para la comparación y la recomendación.
- **Test Items (ítems de test)**: banco de ítems que define estructura, opciones y metadatos de las preguntas. Permite construir cuestionarios de forma modular y validar respuestas de manera uniforme.

En conjunto, estas colecciones permiten: (i) ingestión estructurada de contenido académico y testimonios, (ii) evaluación psicométrica reproducible a partir de reglas declarativas, (iii) enriquecimiento semántico mediante etiquetas y (iv) una capa de presentación desacoplada a través de tarjetas. Esta organización habilita tanto la recuperación semántica (RAG) como el motor de recomendación basado en el alineamiento entre el perfil del usuario y los atributos de las carreras.

7.2.4. Estructuras y modelos de datos vectoriales

Para proporcionar solo la información relevante al LLM y mejorar la rapidez de respuesta, se definieron las siguientes secciones para cada carrera:

1. **Overview:** overview general de la carrera. Incluye los campos:
 - Facultad
 - Duración
 - Empleabilidad
 - Descripción
2. **Development areas:** áreas de desarrollo potencial de la carrera. Incluye los campos:
 - Areas de desarrollo potencial
3. **Study plan:** cursos por año y semestre. Incluye los campos:
 - Año
 - Semestre
 - Cursos
4. **Training areas:** áreas de formación de la carrera. Incluye los campos:
 - Áreas de formación
5. **Graduate profile:** perfil que se espera del egresado de la carrera. Incluye los campos:
 - Perfil del egresado
6. **Competencies:** competencias que se desarrollan en la carrera. Incluye los campos:
 - Competencias desarrolladas
7. **Salary:** rango salarial estimado para un profesional en esta área. Incluye los campos:
 - Salario mínimo
 - Salario máximo
 - Moneda del salario
8. **Career tags:** características relevantes para la carrera. Incluye los campos:
 - Nombre de la etiqueta

- Puntaje de la etiqueta

Estructura y procesamiento de datos vectoriales

Para la estructura de los vectores, se definieron los siguientes campos:

- **id**: identificador único del vector, se compone del id de la carrera en DocumentDB, la sección del vector y el orden de la sección. Ejemplo: `careers:123:overview:1`
- **career_slug**: nombre de la carrera en un formato aceptable para búsquedas semánticas. Ejemplo: `ingenieria_de_sistemas`
- **collection**: colección de la base de datos donde se encuentra la carrera. Ejemplo: `careers`
- **docId**: identificador único del documento en DocumentDB. Ejemplo: `1232...25`
- **order**: orden de la sección del vector. Ejemplo: `1`
- **section**: sección del vector. Ejemplo: `overview`
- **type**: tipo de información que contiene el vector. Ejemplo: `career_overview`

Ventajas de la descomposición de la información en vectores

- **Información fragmentada**: permite fragmentar la información de la carrera en diferentes vectores.
- **Consulta semántica**: permite realizar consultas semánticas más precisas, ya que se puede obtener solo la información relevante para la consulta.
- **Tiempo de respuesta**: reduce el tiempo de respuesta de las consultas, al no obtener toda la información de la carrera.
- **Costo de consulta**: se reduce el costo de la consulta al LLM, ya que se le envía menos tokens por procesar.
- **Costo de actualización**: se reduce el costo de actualización de datos, ya que al momento de actualizar una sección dentro de la carrera, solo se actualiza el vector correspondiente.

Características de la base vectorial seleccionada

Los vectores serán almacenados en una base de datos vectorial hosteada en **Pinecone** (sec. 3).

- **Dimensionalidad (sec. 5.17.6)**: se seleccionó una dimensionalidad de **768**.
- **Tipo de vector (sec. 5.17.6)**: se seleccionó un **vector denso**.
- **Función de los embeddings (sec. 5.17.6)**: se seleccionó la función **cosine similarity** para la función de búsqueda de embeddings.

7.2.5. Pruebas funcionales

Con el objetivo de garantizar la consistencia y fiabilidad de la información almacenada en la base de conocimiento del sistema, se propone la realización de las siguientes pruebas: **(i) pruebas automáticas de integridad estructural** y **(ii) pruebas de precisión de consulta**. Estas evaluaciones permitirán verificar que los documentos almacenados en la base NoSQL (DocumentDB) cumplan con la estructura esperada y que los datos puedan ser recuperados correctamente mediante consultas controladas.

Pruebas automáticas de integridad (marco de prueba del objetivo 1 - 3.2)

Estas pruebas tendrán como propósito confirmar que todos los registros de carreras universitarias incluyan los campos esenciales definidos durante el diseño de la base de conocimiento especificados. Cada documento debe contener los atributos:

- nombre de la carrera
- facultad a la que pertenece
- duración en años
- plan de estudio
- perfil del egresado
- empleabilidad
- salario mínimo
- salario máximo

los cuales permiten tanto la identificación de la carrera como su indexación dentro del sistema de recomendación y guianza vocacional.

Para automatizar el proceso de validación se propone la realización de un script en *Python*, utilizando el conector oficial de *MongoDB*. El script recorrerá cada documento de la colección **carreras** y verificará la existencia de los campos requeridos mediante una sentencia de validación booleana.

Los resultados de esta prueba se almacenarán en un registro de validación, en el cual se contabilizará la cantidad de registros incompletos y se verificará que el porcentaje de cumplimiento de la estructura sea igual o superior al **95 %**, valor considerado como el umbral mínimo de calidad de la información en bases documentales. **Acorde a los requisitos establecidos en el objetivo específico 1 (sec. 3.2)**

La integridad se calculará según la ecuación:

$$\text{Integridad} = \frac{\text{Documentos válidos}}{\text{Documentos totales}} \times 100 \quad (1)$$

El cumplimiento de este umbral asegurará que la base de conocimiento mantenga una estructura homogénea y apta para las consultas del sistema RAG y del modelo de recomendación.

Pruebas de precisión de consulta (marco de prueba del objetivo 2 - 3.2)

Una vez validada la estructura interna de los documentos, se evaluará la **precisión de recuperación de datos** mediante un conjunto de 20 consultas de control. Estas consultas simularán preguntas típicas que el sistema debe responder correctamente, tales como:

- ¿Qué facultad ofrece la carrera de Ingeniería Química?
- ¿Cuál es la duración de la carrera de Psicología?
- ¿Se lleva Base de datos 1 en tercer año, primer semestre, de la carrera de Ingeniería en Ciencia de la Computación?

Cada consulta se ejecutará directamente sobre la base de datos a través de la API interna, y se comparará el resultado obtenido con la respuesta esperada según la información original de las carreras.

El desempeño se medirá utilizando la métrica de **precisión** (P), definida como:

$$P = \frac{\text{Consultas correctas}}{\text{Consultas totales}} \times 100 \quad (2)$$

Se considerará como *consulta correcta* aquella en la que la respuesta devuelta coincida completamente con la información esperada, sin errores de valor ni de formato. Los resultados obtenidos se tabularán, mostrando el número de consultas exitosas y el porcentaje de precisión general.

El criterio de aceptación establecido será alcanzar una precisión mínima del **95 %**, lo que garantizará la confiabilidad del sistema para recuperar información académica dentro de los procesos de orientación vocacional. **Acorde a los requisitos establecidos en el objetivo específico 2 (sec. 3.2)**

7.2.6. Resultados de la prueba

A continuación se presenta los resultados de la ejecución de la prueba de precisión de consulta y la prueba de integridad estructural, los resultados de la ejecución de la prueba se encuentran en el anexo 29.

Resultados de integridad estructural

Cuadro 13. Resultados de la prueba de integridad estructural

[1] Integridad estructural	
- Documentos totales	: 33
- Documentos válidos	: 33
- Precisión (integridad)	: 100.0%

La primera fase de validación se centró en la integridad estructural de la base de conocimiento. Como se detalla en el Cuadro 13, la prueba automática verificó los 33 documentos de carreras, encontrando que todos eran estructuralmente válidos y completos.

Resultados de precisión de consulta

Cuadro 14. Resultados de la prueba de precisión de consulta

[2] Precisión de consulta (Q/A controladas)	
- Pruebas totales	: 30
- Pruebas correctas	: 30
- Precisión (consultas)	: 100.0%

Una vez verificada la integridad, se procedió a evaluar la exactitud del contenido mediante 30 consultas controladas. Como muestra el Cuadro 14, el sistema recuperó la información correcta en la totalidad de los casos, alcanzando un nivel de precisión del 100.0 %.

7.3. Módulo de guianza vocacional (obj. específicos 3 y 4 - 3.2)

Con el objetivo de brindar a los estudiantes una herramienta interactiva de orientación vocacional y académica, se propone la implementación de un sistema basado en el enfoque **Retrieval-Augmented Generation (RAG sec. 5.15.3)**. Este sistema tiene como finalidad responder preguntas relacionadas con carreras universitarias, planes de estudio y contenidos temáticos, utilizando información confiable y actualizada proveniente de las fuentes institucionales.

La metodología contempla el uso de un **modelo de lenguaje preentrenado (LLM)** para el razonamiento y la generación de respuestas naturales (véase **Sec. 5.2.3**), integrado con un **Retrieval-Augmented Generation (RAG)** que utiliza una **base de datos vectorial (Vector DB)** encargada de la búsqueda semántica y recuperación de contexto (véase **Sec. 5.17.6**) y una **base de datos NoSQL por Documentos** que almacena la información estructurada sobre carreras, cursos y pensum académico (véase **Sec. 5.17.7**).

7.3.1. Justificación de elecciones técnicas

Justificación de la elección de RAG

- **Actualización:** se puede actualizar la información sin la necesidad de reentrenar al modelo para aplicar los conocimientos.
- **Frescura:** la información proviene de fuentes externas confiables, asegurando una respuesta actual y verificable sin reentrenamiento.
- **Modificación:** no modifica pesos del modelo, sino que recupera documentos relevantes para la pregunta.
- **Costos:** bajo entrenamiento, pero requiere infraestructura de recuperación para obtener la información relevante.

En consecuencia, se seleccionó RAG por su capacidad de actualización dinámica de la información, sin necesidad de reentrenar el modelo, y su bajo costo de entrenamiento.

Justificación de la elección de Google - Gemini embeddings

- **Rendimiento:** su rendimiento es muy superior a los otros modelos de embeddings, hasta el momento de la redacción de este trabajo.
- **Facilidad de uso:** es muy sencillo de integrar dentro de un sistema, ya que se puede utilizar el SDK de Google para Python, JavaScript, Go, entre otros.
- **Dimensionalidad:** se puede configurar para cualquier cantidad de dimensiones que requiramos, siempre y cuando sea menor a 2048 dimensiones.
- **Costo:** el precio es un poco más elevado comparado con los demás modelos, sin embargo, considerando la diferencia de rendimiento, es un precio razonable.

En consecuencia, se seleccionó Google - Gemini Embeddings como modelo de embeddings, debido a su rendimiento superior, facilidad de uso y dimensionalidad configurable.

Justificación de la elección de Gemini 2.0 Flash

- **Rendimiento:** posee un rendimiento muy equilibrado, siendo multilinguaje y multi-modal.
- **Velocidad:** es muy rápido, lo que lo hace ideal para tareas en tiempo real.
- **Costo:** el precio es muy accesible, siendo el más económico de los modelos de LLMs.
- **Plan gratuito:** posee un plan gratuito muy generoso, siendo el más generoso de los modelos de LLMs.

- **Integración:** posee una integración muy sencilla, gracias a su gran variedad de SDKs disponibles y documentación muy completa.

En consecuencia, se seleccionó Gemini 2.0 Flash como modelo de LLMs, debido a su rendimiento equilibrado, velocidad, costo accesible, plan gratuito generoso y buena integración.

7.3.2. Desarrollo

Arquitectura del sistema de guianza vocacional

La arquitectura propuesta responde a las necesidades de proveer respuestas contextualizadas, precisas y actualizadas, sin requerir reentrenar el modelo base. Por esta razón, se adoptó un enfoque de **Retrieval-Augmented Generation (RAG sec. 4)**, en el cual se utilizará el LLM de **Gemini-2.0-Flash (sec. 6)** para generar respuestas a partir de información recuperada dinámicamente desde la base de datos de conocimiento.

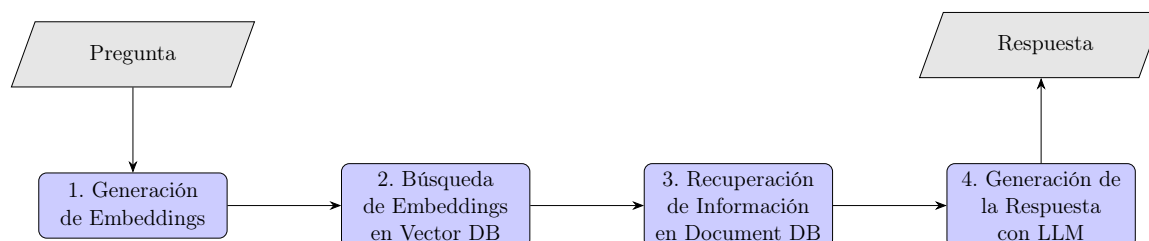
Para el desarrollo del sistema de guianza vocacional, se propone la siguiente arquitectura:

- **Componente de generación de embeddings:** componente encargado de generar los embeddings de las preguntas del usuario.
- **Componente de recuperación:** componente de recuperación encargada de recuperar la información de la base de datos de información, utilizando los índices obtenidos de los vectores.
- **Componente de generación:** componente de generación de la respuesta, encargado de generar la respuesta natural para el usuario, utilizando el contexto de la información relevante. Aquí se hace uso de **Prompt Engineering (sec. 5.6.3)** para agregar el contexto necesario a la pregunta del usuario.

Aquí se cumple el **objetivo específico 3 (sec. 3.2)** de implementar un sistema de guianza vocacional basado en un Modelo de Lenguaje Extenso (LLM) que utilice la base de conocimiento construida para responder preguntas sobre carreras universitarias.

Componentes principales del sistema de guianza vocacional

Figura 22. Diagrama de la arquitectura del sistema de guianza vocacional



1. **Componente de generación de embeddings:** se encarga de generar los embeddings de las preguntas del usuario, utilizando el modelo de embeddings de **Google - gemini-embedding-001 (sec. 5)** para generar los embeddings de las preguntas del usuario. Este componente es responsable de transformar las preguntas del usuario en vectores, para luego ser utilizados en la búsqueda semántica de la información relevante.
2. **Componente de búsqueda de embeddings en Vector DB:** se encarga de buscar los embeddings de las preguntas del usuario en la base de datos de vectores. Para obtener los índices de la información relevante de la base de datos de información académica, para ser recuperados de Document DB.
3. **Componente de recuperación de la información en Document DB:** se encarga de recuperar la información de la base de datos de información académica, utilizando los índices obtenidos de los vectores. Para poder ser utilizados como contexto para la generación de la respuesta.
4. **Componente de generación de la respuesta con LLM:** se encarga de generar la respuesta natural para el usuario, utilizando el contexto de la información relevante. Esto permite que el modelo genere respuestas solo con la información relevante y actualizada, sin necesidad de memorizar la información.

Consideraciones éticas y de seguridad

Dado que el sistema manejará datos de carácter educativo, se establecerán medidas específicas para proteger la privacidad, garantizar la transparencia y reducir riesgos asociados al uso de modelos de lenguaje. Estas medidas abarcan el control de acceso a la base de datos, la protección de los datos en tránsito y en reposo, así como la aplicación de filtros sobre las preguntas y respuestas.

Restricciones de acceso a la base de datos.

El LLM no accederá directamente a la base de datos, sino a través de un módulo intermedio encargado de extraer lo más importante de la pregunta y obtener la información relevante de la base de datos. Este diseño evita que el modelo pueda ejecutar comandos arbitrarios o modificar los datos, en caso el se haya utilizado un prompt malicioso, y de esta forma solo se obtendrá la información mínima necesaria para responder a cada pregunta.

Filtros en las preguntas y respuestas.

Para garantizar que el modelo se limite al dominio educativo y vocacional, se incluirán filtros previos y posteriores a la generación de respuestas. Los filtros previos analizarán las preguntas de los estudiantes, bloqueando aquellas que contengan solicitudes fuera del dominio definido (por ejemplo, consultas personales, inapropiadas o que intenten inyectar código). Por su parte, los filtros posteriores revisarán las respuestas generadas para detectar lenguaje ofensivo, sesgos o información inconsistente con los datos oficiales. Este doble filtrado reducirá riesgos de uso indebido y mejorará la calidad de la interacción con el sistema.

En conjunto, estas medidas buscan asegurar que el sistema de orientación vocacional se despliegue en un entorno confiable, ético y seguro, en el que tanto la información como la experiencia de usuario estén protegidas frente a vulnerabilidades técnicas y riesgos sociales.

7.3.3. Pruebas funcionales

Con el objetivo de garantizar la consistencia y fiabilidad de las respuestas del sistema, se propone la realización de la siguiente prueba que evaluará: **(i) exactitud factual**, **(ii) relevancia semántica** y **(iii) claridad y completitud**. Estas evaluaciones permitirán verificar que las respuestas del sistema sean precisas y diversas.

Criterios de evaluación

Se establecerán tres criterios principales para medir el desempeño del sistema de guianza:

- **Exactitud factual:** grado en el que las respuestas contengan datos correctos y verificables según la información disponible en la base de conocimiento.
- **Relevancia semántica:** nivel de correspondencia entre la pregunta realizada y la información presentada por el sistema.
- **Claridad y completitud:** medida en la que la respuesta sea comprensible, estructurada y suficiente para satisfacer la consulta del usuario.

Diseño de la prueba (marco de prueba del objetivo 4 - 3.2)

Para la validación, se propone la elaboración de un **conjunto de 20 preguntas de prueba** relacionadas con carreras universitarias, cursos, duración de programas y áreas de desarrollo profesional. Estas preguntas fueron diseñadas a partir de las categorías más frecuentes en los procesos de orientación vocacional. Ejemplos de las consultas utilizadas serán:

- ¿Cuántos años dura la carrera de Ingeniería Mecánica?
- ¿Qué facultad ofrece la carrera de Psicología Industrial?
- ¿Cuáles son las principales áreas de desarrollo de Diseño Gráfico?

Cada pregunta será procesada a través del sistema de guianza vocacional, el cual utilizará el modelo RAG para recuperar la información relevante desde la base vectorial y generar una respuesta textual mediante el LLM.

Evaluación por jueces

Las respuestas obtenidas serán evaluadas por un grupo de **tres jueces humanos**, quienes calificarán cada una según los tres criterios definidos. Se utilizará una **escala ordinal de 0 a 5**, donde 0 representa un desempeño nulo y 5 un desempeño excelente.

Cada criterio se puntuará de forma independiente, y posteriormente se calculará un promedio global por pregunta.

$$\text{Puntaje promedio} = \frac{\text{Puntaje factual} + \text{Puntaje relevancia} + \text{Puntaje claridad}}{3} \quad (3)$$

Para obtener el índice global de fidelidad del sistema, las puntuaciones promedio se normalizarán en una escala porcentual de 0 a 100, según la ecuación:

$$\text{Fidelidad} = \frac{\text{Promedio de puntaje obtenido}}{5} \times 100 \quad (4)$$

Criterio de aceptación

Se establecerá como valor mínimo aceptable una fidelidad del **85 %**, lo cual representa un desempeño alto en la precisión y pertinencia de las respuestas. Este umbral fue determinado en función del **objetivo específico 4 (sec. 3.2)** de alcanzar una fidelidad de la información superior al 85 % en un conjunto de prueba controladas, y de los estándares de calidad para sistemas de recuperación de información educativa y modelos de lenguaje aplicados a entornos académicos.

7.3.4. Resultados de la prueba

A continuación, se presentan los resultados de la prueba realizada a 20 preguntas relacionadas con carreras universitarias, cursos, duración de programas y áreas de desarrollo profesional. Las preguntas de la prueba y sus métricas se encuentran en el anexo 11.1.3, y la captura completa de los resultados de la prueba se encuentran en la figura 35.

Métricas generales de la prueba

Cuadro 15. Resultados generales de la prueba

Métrica	Valor
Exactitud	4.92
Relevancia	4.85
Claridad	4.9
Puntaje promedio	4.89
Fidelidad	97.78 %

Al ver los resultados generales (Cuadro 15) se puede ver que el sistema tiene un buen desempeño en la prueba con los siguientes puntales:

- **Exactitud factual (4.92):** significa que el modelo utilizó información recuperada de la base de conocimiento, no su propio conocimiento, para generar la respuesta.

- **Relevancia semántica (4.85):** significa que la respuesta del modelo es relevante para la pregunta realizada.
- **Claridad y completitud (4.9):** significa que la respuesta del modelo es clara y completa, al momento de responder a la pregunta del usuario.
- **Fidelidad (97.78 %):** nos indica que el modelo utilizó muy bien la información recuperada para sustentar su respuesta, no utilizando su propio conocimiento, manteniendo una buena consistencia y relevancia en cada respuesta.

Aunque si se vió una disminución en la puntuación considerable en las preguntas **19 (Cuadro 38)** y **20 (Cuadro 39)**, probablemente porque el modelo no pudo recuperar la información relevante del historial del consultas, necesitando que se le indicara nuevamente las carreras para contestar las preguntas.

7.4. Módulo de recomendación (obj. específicos 5, 6 y 7 - 3.2)

Este módulo se encargará de la recomendación de carreras universitarias y contenidos basado en las características de los usuarios. Para ello, se propone la implementación de un sistema de recomendación que pueda analizar los datos de los usuarios y recomendar tanto las carreras universitarias como el contenido temático con las características de los usuarios extraídas de una prueba de evaluación.

7.4.1. Justificación de elecciones técnicas

Justificación de la elección de Python

- **Experiencia:** Python es el lenguaje de programación que se ha utilizado en varios ejercicios y proyectos de la carrera, lo que hace que sea el lenguaje más familiar para el desarrollo de modelos de Machine Learning.
- **Madurez:** Python es el lenguaje de programación más utilizada en el sector de Machine Learning, lo que lo hace el lenguaje con mayor soporte y documentación para cualquier uso, en especial para Machine Learning.
- **Soporte:** Python tiene una gran cantidad de librerías y herramientas para una gran variedad de modelos y algoritmos de Machine Learning, lo que facilita la implementación de cualquier solución de Machine Learning.
- **Velocidad:** por ser un lenguaje interpretado, Python tiende a tener un rendimiento inferior a los lenguajes compilados, sin embargo, con la ayuda de librerías optimizadas y la habilitación del uso de GPU, se puede llegar a tener un rendimiento aceptable.

En consecuencia, se seleccionó Python por la experiencia previa del investigador en el lenguaje, la madurez y el soporte para Machine Learning que goza el lenguaje, y las mejoras en el rendimiento que se han obtenido gracias al uso de librerías optimizadas y la habilitación del uso de GPU.

Justificación de la elección de árboles de decisión

- **Complejidad computacional:** media, lo que lo hace un algoritmo adecuado para su ejecución dentro de un ambiente de producción, permitiéndonos mejorar el tiempo de respuesta del sistema.
- **Requerimiento de datos:** moderado, lo que hace que no se requiera de una cantidad excesiva de datos para su entrenamiento, lo cual es ideal para el caso de estudio.
- **Paradigma:** supervisado, nos ayuda a tener un modelo de machine learning que sea más fácil de interpretar y entender, lo cual faltaría la explicación y justificación de las recomendaciones del sistema.

En consecuencia, se seleccionó Árboles de Decisión por su complejidad computacional, requerimiento de datos, paradigma y la facilidad de interpretación y entendimiento del modelo.

Justificación de la elección de gradient boosting (GBT / XGBoost)

1. **Capacidad de ranking explícita:** está específicamente diseñado para tareas de ranking.
2. **Alta precisión en clasificación multiclase:** tiene el poder predictivo más alto entre los algoritmos comparados, lo que es crítico para este problema de clasificación multiclase.
3. **Manejo de datos desbalanceados:** maneja explícitamente este tipo de datos desbalanceados, ajustando automáticamente los pesos de las carreras sub-representadas para que no sean ignoradas por el modelo.
4. **Flexibilidad para capturar relaciones complejas:** al entrenar árboles de forma secuencial y corrigiendo errores previos, puede capturar estas relaciones sutiles y complejas mejor que árboles de decisión únicos.
5. **Control del sobreajuste mediante regularización:** incluye mecanismos de regularización integrados que pueden ajustarse durante el entrenamiento.

En consecuencia, se seleccionó Gradient Boosting (GBT / XGBoost) por su capacidad de ranking explícita, alta precisión en clasificación multiclase, manejo de datos desbalanceados, flexibilidad para capturar relaciones complejas y control del sobreajuste mediante regularización.

7.4.2. Desarrollo

Creación del conjunto de datos

En esta sección se define la recolección del conjunto de datos inicial para poder entrenar el modelo de recomendación. Esta recolección se realizará mediante la aplicación de una prueba de evaluación a un grupo de usuarios y la recopilación de los resultados.

Definición del conjunto de datos

Para la creación del **conjunto de datos (o dataset sec. 5.16.3)**, el primer paso es definir las **características (sec. 5.16.5)** que se utilizarán para que el **modelo (sec. 5.3.2)** identifique los perfiles. En esta etapa, se deben seleccionar las variables numéricas que describan con mayor precisión a una persona. El objetivo es aplicar técnicas de **ingeniería de características (sec 5.16.6)** para transformar los datos y así establecer la compatibilidad entre los perfiles de los usuarios y los requerimientos de las carreras universitarias.

Antecedentes y justificación de la selección de instrumentos

La elección de los **instrumentos de evaluación (sec. 5.6.5)** incluidos en esta investigación se sustentó en estudios previos de carácter internacional, que han demostrado su validez y fiabilidad en la medición de **competencias cognitivas (sec. 5.9.1)** y **no cognitivas (sec. 5.9.2)** en diferentes **poblaciones (sec. 5.5)**.

Por un lado, el **Programme for the International Assessment of Adult Competencies (PIAAC) (sec. 5.10.1)** de la OCDE ha sido ampliamente utilizado para evaluar habilidades clave como comprensión lectora, matemáticas y resolución de problemas en entornos tecnológicos. Según el estudio de [173], publicado en Large-scale Assessments in Education, los **ítems (sec. 5.6.4)** del PIAAC cuentan con un marco metodológico robusto que asegura la validez intercultural y la comparabilidad de los resultados entre contextos diversos [173]. Esto permite que las métricas derivadas de dichos ítems sean relevantes para investigaciones que buscan evaluar competencias transversales aplicables a distintos entornos académicos y profesionales.

Asimismo, documentos técnicos recientes de la [174] resaltan la importancia de adaptar las evaluaciones a escenarios de aprendizaje y **orientación vocacional (sec. 5.11.4)**, donde la validez de constructo y la fiabilidad psicométrica resultan fundamentales para la toma de decisiones educativas [174]. En este sentido, la inclusión de ítems basados en estas pruebas dentro de la presente investigación asegura que las características extraídas de los participantes reflejen de manera precisa sus habilidades cognitivas y capacidades de resolución adaptativa de problemas, aspectos críticos para la recomendación de trayectorias educativas.

En el marco de este trabajo, los ítems derivados de estos estudios se integran como base para la construcción de las características de los usuarios, de forma que los resultados obtenidos no solo se alinean con estándares internacionales, sino que también permiten generar un modelo de recomendación vocacional sólido y comparable con investigaciones previas en el área.

Instrumentos seleccionados Basándonos en los antecedentes mencionados, se seleccionaron los siguientes instrumentos psicométricos para evaluar las características de los usuarios:

- **Test de personalidad (big five):** basado en el **modelo de los cinco grandes rasgos de personalidad (sec. 5.7.1)**, este instrumento utiliza ítems diseñados para cuantificar numéricamente cada uno de los rasgos en los individuos.
- **Prueba de habilidades cognitivas (PIAAC):** se desarrolló una versión reducida

del Programa para la Evaluación Internacional de Competencias de Adultos. Esta prueba evalúa competencias clave como comprensión lectora, razonamiento matemático y resolución de problemas en personas mayores de 16 años.

- **Prueba de habilidades académicas (PAA):** corresponde a una versión abreviada de la **Prueba de Aptitud Académica (sec. 5.10.2)**, diseñada para medir el razonamiento verbal y matemático en estudiantes de bachillerato.
- **Test de intereses vocacionales (RIASEC):** fundamentado en la tipología de la **teoría de Holland (sec. 5.7.2)**, los ítems seleccionados permiten obtener una puntuación numérica para cada una de las seis áreas de interés vocacional.
- **Test de persistencia y consistencia (Grits):** basada en la **teoría del Grit (sec. 5.7.3)**, esta escala cuantifica la perseverancia y la pasión de un individuo hacia sus metas a largo plazo.

Definición de los ítems Dado que el tiempo de los usuarios es limitado, no es factible aplicar la totalidad de los ítems de cada instrumento de evaluación. Por lo tanto, se realizó una selección estratégica de ítems para cada prueba. El objetivo es mantener la capacidad de cada instrumento para capturar la información relevante del usuario, pero reduciendo significativamente el tiempo de aplicación. Este proceso de selección fue supervisado por la psicóloga del proyecto para garantizar la validez y fiabilidad de las mediciones. Los ítems se encuentran descritos a profundidad en el anexo 11.1.1.

Recolección de los datos

Esta fase se centró en la recopilación de datos de estudiantes universitarios activos. Se administró la prueba con los instrumentos de evaluación posteriormente seleccionados y se vincularon los perfiles resultantes de cada participante a la carrera universitaria que cursaban. Durante el muestreo, se vigiló la obtención de un número de casos suficiente para garantizar la representatividad estadística de cada carrera. El objetivo final fue consolidar un conjunto de datos donde los perfiles de competencias, personalidad e intereses sirvan como descriptores para cada una de las titulaciones académicas.

Participantes y estrategia de muestreo La población objetivo del estudio fueron estudiantes universitarios que se encontraban cursando activamente cualquier carrera de nivel de pregrado. Se empleó un muestreo no probabilístico por conveniencia y de bola de nieve. Inicialmente, se contactó a una red de estudiantes accesibles y se les invitó a participar. Posteriormente, se les animó a compartir el enlace del estudio con otros compañeros de sus respectivas carreras.

Plataforma de recolección Para facilitar y estandarizar la recolección de los datos, se utilizó la herramienta en línea **PsyToolkit (sec. 5.17.4)**. Esta plataforma permitió centralizar la administración de los instrumentos de evaluación seleccionados y obtener los resultados directamente en un formato unificado, lo que simplificó significativamente las fases de recopilación y análisis posterior.

Proceso de recolección de datos El procedimiento seguido por cada participante fue el siguiente:

1. **Acceso a la plataforma:** los participantes accedieron a la batería de pruebas a través de un enlace web único.
2. **Consentimiento informado:** antes de iniciar, se presentó un formulario de consentimiento informado. Este documento detallaba los objetivos del estudio, la naturaleza voluntaria de su participación, las medidas para garantizar el anonimato y el uso estrictamente académico de la información. La aceptación explícita de estos términos fue un requisito indispensable para continuar.
3. **Registro de carrera universitaria:** seguidamente, se solicitó al participante que indicara su sexo y la carrera universitaria que estaba cursando. Estos fueron los únicos datos demográficos y de categorización recolectados, siendo esencial para el etiquetado de los perfiles resultantes.
4. **Administración de las pruebas:** el participante procedió a completar el conjunto de instrumentos de evaluación. La plataforma registró las respuestas de forma automática.
5. **Finalización y envío:** al concluir, las respuestas fueron enviadas y almacenadas de forma anónima en la base de datos centralizada para su posterior análisis.

Privacidad, confidencialidad y consideraciones éticas Se implementó un protocolo estricto para proteger la identidad y los datos de los participantes.

1. **Consentimiento y propósito:** al inicio de la prueba, se le indicó a cada participante el propósito académico de la recopilación y se aseguró que sus resultados no serían utilizados para fines comerciales o ajenos a los objetivos de este trabajo de graduación.
2. **Anonimato total:** no se solicitó ninguna clase de información de identificación personal, como nombre, correo electrónico, país de residencia, nivel socioeconómico, etc. Esto se hizo con el propósito de proteger la confidencialidad y garantizar que los resultados no pudieran ser vinculados de vuelta a los participantes.
3. **Uso restringido de los datos:** los resultados finales no serán colocados en disponibilidad pública, incluso después de terminado el trabajo profesional, sino que serán exclusiva y estrictamente utilizados dentro del margen de entrenamiento del modelo. Para cualquier publicación o reporte, solo se utilizarán datos agregados.
4. **Seguridad y eliminación de datos:** los resultados se mantendrán en un repositorio seguro durante la duración del estudio y serán eliminados de manera permanente una vez que el proyecto haya concluido y todos los análisis necesarios hayan sido realizados.

Prueba de concepto

A continuación, se presenta la prueba de concepto realizada para validar la implementación del sistema de recomendación. Se presentan los resultados obtenidos y las limitaciones encontradas; así como la justificación. Esto cumple parcialmente con el **objetivo específico 6 (sec. 3.2)**, de generar perfiles de usuario multidimensionales mediante la aplicación de una batería de pruebas psicométricas seleccionadas para medir rasgos determinantes de un perfil.

Cantidad de participantes necesaria Para poder tener una muestra representativa, se necesitará un **mínimo de 20 resultados de estudiantes por carrera**. Es decir, se necesitará un mínimo de $20 \times 33 \text{ carreras universitarias} = 660 \text{ resultados de estudiantes}$. Esto ya que generar el perfil de cada carrera es un proceso complejo, que requiere de una gran cantidad de datos para poder crearlo con suficiente precisión.

Obstáculos y limitaciones

- **Resultados necesarios por carrera:** se necesita de una cantidad considerable de resultados para poder crear el perfil de cada carrera con suficiente precisión. Teniendo en cuenta la complejidad de la recomendación, el mínimo de 20 resultados es incluso un poco insuficiente, más adecuado para poder obtener resultados con suficiente precisión.
- **Cantidad de carreras:** la cantidad de carreras universitarias disponibles dificulta la recolección de datos. Se tiene un total de **33 carreras universitarias** disponibles. Lo que hace que la recolección de datos sea más compleja, ya que se necesita recolectar de cada carrera una cantidad considerable de resultados para el perfil.
- **Duración de la prueba:** la prueba de evaluación es una prueba larga, que puede tomar de 10 a 30 minutos para ser completada. Esto dificulta la recolección de datos, ya que se necesita un tiempo considerable para que los estudiantes completen la prueba.
- **Necesidad de resultados completos:** por la naturaleza de la prueba y el objetivo de la misma, se necesita que los estudiantes completen la prueba. Esto dificulta la recolección de datos, ya que cualquier resultado parcial o incompleto no será útil para el entrenamiento del modelo.

Resultados recolectados en total

Figura 23. Participantes del test hecho en la plataforma PsyToolkit

Number of datafiles/participants	
Total visitors to welcome screen	291
Completed surveys	72
Uncompleted surveys	61
Total surveys (complete or incomplete)	133

La recolección se llevó desde el 9 de septiembre hasta el 25 de octubre del 2025, con ayuda de los directores de carrera de las distintas facultades. Sin embargo, por las limitaciones de tiempo, y los obstáculos y limitaciones mencionados previamente, se logró llegar solamente a **291 participantes**; de ellos el **45.7 % (133 participantes)** siguieron luego del consentimiento informado, para terminar con un **45.8 % (61)** de las pruebas completadas y con el restante **54.2 % (72)** de las pruebas incompletas.

Cantidad recolectada por carrera

Figura 24. Cantidad de participantes recolectados por carrera

Item	Score	Count	Percentage
Ingeniería Biomédica	1	5	5.0
Ingeniería Biotecnología Industrial	2	na	na
Ingeniería Ciencia de la Administración	3	2	2.0
Ingeniería Ciencias de Alimentos	4	na	na
Ingeniería Ciencias de Alimentos Industrial	5	na	na
Ingeniería Civil	6	6	6.0
Ingeniería Civil Arquitectónica	7	3	3.0
Ingeniería Ciencia de la Computación y Tecnologías de la Información	8	47	47.0
Ingeniería Sistemas de Información Computacional	9	na	na
Ingeniería Electrónica	10	4	4.0
Ingeniería Industrial	11	na	na
Ingeniería Mecánica	12	na	na
Ingeniería Mecánica Industrial	13	na	na
Ingeniería Mecatrónica	14	8	8.0
Ingeniería Química	15	1	1.0
Ingeniería Química Industrial	16	na	na
Ingeniería en Ciencia de la Administración	17	5	5.0
Administración de Empresas	18	6	6.0
International Marketing and Business Analytics	19	na	na
Comunicación Estratégica	20	na	na
Profesorado de Enseñanza Media especializado en Educación Musical	21	na	na
Profesorado de Enseñanza Media especializado en English Language Teaching (ELT)	22	na	na
Profesorado Especializado en Educación Inclusiva	23	na	na
Profesorado Especializado en Educación Primaria (100% virtual)	24	na	na
Profesorado Especializado en Problemas del Aprendizaje	25	na	na
Profesorado de Enseñanza Media Especializado en Matemática y Ciencias Físicas	26	na	na
Profesorado de Enseñanza Media Especializado en Ciencias Químicas y Biológicas	27	na	na
Profesorado de Enseñanza Media Especializado en Ciencias Sociales	28	na	na
Profesorado de Enseñanza Media Especializado en Comunicación y Lenguaje	29	na	na
Baccalaureatus en Artibus	30	na	na
Baccalaureatus en Scientiis	31	na	na
Arquitectura	32	na	na
Biología	33	1	1.0
Bioquímica y Microbiología	34	1	1.0
Biotecnología Molecular	35	1	1.0
Física	36	na	na
Matemática Aplicada	37	na	na
Nutrición	38	na	na
Química	39	na	na
Química Farmacéutica	40	na	na
Antropología	41	na	na
Arqueología	42	na	na
Psicología	43	1	1.0
Relaciones Internacionales	44	na	na
Composición y Producción Musical	45	8	8.0
Diseño de Producto e Innovación	46	1	1.0

Al analizar la cantidad recolectada por carrera, podemos ver que todas las carreras, exceptuando la carrera de **Ingeniería en Ciencias de la Computación y Tecnologías de la Información** la cuál se registró el **47 %** de los participantes, tienen una cantidad de participantes menor a 10; lo cuál era la mitad de lo que se necesitaba para tener datos suficientes para el entrenamiento del modelo.

Con ello, terminamos con un total de **61 resultados de estudiantes utilizables** para el entrenamiento del modelo de machine learning, y solo una carrera de las 33 disponibles, **Ingeniería en Ciencias de la Computación y Tecnologías de la Información**, lo que no es suficiente para el entrenamiento del modelo.

Alternativa de solución

Para poder tener datos suficientes para el entrenamiento del modelo, se propone la generación de datos sintéticos para poder completar la cantidad de datos necesarios para el entrenamiento del modelo. Esto se hará mediante generando valores aleatorios para las características de las carreras y los perfiles de los estudiantes dentro del rango correspondiente a cada característica.

Archivos creados para el entrenamiento Con los datos de los usuarios generados, se crearán 3 distintos archivos:

- **users.csv:** archivo con los resultados de cada participante en cada una de las pruebas

realizadas.

- **careers.csv:** archivo con los resultados del perfil de cada carrera, generado de la combinación de los resultados de los estudiantes pertenecientes a la carrera.
- **training_pairs.csv:** archivo creado por la naturaleza del algoritmo de machine learning a utilizar, que posee una fila por cada perfil de carrera dentro de las demás carreras disponibles. Los valores se encuentran en rangos bajos para establecer enfatizar la carrera a la que pertenece el perfil.

Selección del algoritmo de machine learning

Esta fase se centra en la construcción del motor predictivo del sistema.

- **Selección del algoritmo:** para este problema de clasificación multiclase (predecir una carrera entre múltiples opciones), se seleccionará un algoritmo de alto rendimiento. El candidato principal es XGBoost (Extreme Gradient Boosting), debido a su excepcional poder predictivo y alta eficiencia computacional. Este modelo es un estándar en la industria para datos tabulares gracias a su robusto sistema de regularización que controla el sobreajuste y su capacidad para capturar relaciones complejas en los datos. Se considerarán otros modelos como Random Forest o Redes Neuronales para realizar análisis comparativos de rendimiento.
- **Ajuste de hiperparámetros:** se utilizarán técnicas como Grid Search o Randomized Search con **validación cruzada** (sec. 5.3.7) para encontrar la configuración óptima de parámetros (ej. el número de árboles en Random Forest) que maximice el rendimiento predictivo [33], [34].
- **Función de pérdida (error):** el proceso de entrenamiento buscará minimizar una función de pérdida. Para la clasificación multiclase, se utilizará la **entropía cruzada categórica** (sec. 5.3.5) [30], [31].

Entrenamiento del modelo

Pasos previos al entrenamiento Para poder entrenar el modelo, ya que no poseemos datos suficientes para el entrenamiento del modelo, se debe de generar datos sintéticos para poder completar la cantidad de datos necesarios para el entrenamiento del modelo. Esto se hará mediante generando valores aleatorios para las características de las carreras y los perfiles de los estudiantes dentro del rango correspondiente a cada característica.

Archivos creados para el entrenamiento Con los datos de los usuarios generados, se crearán 3 distintos archivos:

- **users.csv:** archivo con los resultados de cada participante en cada una de las pruebas realizadas.
- **careers.csv:** archivo con los resultados del perfil de cada carrera, generado de la combinación de los resultados de los estudiantes pertenecientes a la carrera.

- **training_pairs.csv:** archivo creado por la naturaleza del algoritmo de machine learning a utilizar, que posee una fila por cada perfil de carrera dentro de las demás carreras disponibles. Los valores se encuentran en rangos bajos para establecer enfatizar la carrera a la que pertenece el perfil.

Entrenamiento del modelo

Con los archivos creados, se procede a entrenar el modelo de machine learning. Para ello, se utilizará el algoritmo de machine learning seleccionado (**Árboles de Decisión**), con su configuración óptima de hiperparámetros, y se entrenará utilizando exclusivamente los archivos creados.

Pasos del entrenamiento:

1. **Presentación de los datos:** se alimenta al modelo con el archivo **training_pairs.csv**, que contiene múltiples ejemplos de perfiles de estudiantes emparejados con carreras universitarias. Cada ejemplo incluye características del estudiante (resultados de las pruebas) y la carrera universitaria asociada.
2. **Aprendizaje de patrones:** el modelo analiza estos ejemplos y busca identificar patrones y relaciones. Por ejemplo, puede identificar que “estudiantes con altas puntuaciones en razonamiento matemático tienden a ser recomendados para carreras de Ingeniería” o que “estudiantes con alta creatividad y apertura a la experiencia tienden a ser recomendados para carreras artísticas”.
3. **Ajuste iterativo:** el modelo ajusta su aprendizaje iterativamente, mejorando gradualmente su capacidad para hacer recomendaciones precisas. En cada iteración, el modelo compara sus predicciones con las respuestas correctas y realiza pequeños ajustes para mejorar.
4. **Validación del aprendizaje:** durante el entrenamiento, se reserva una parte de los datos (validación cruzada) para verificar que el modelo no solo memoriza los datos, sino que realmente ha aprendido patrones generalizables. Esto asegura que el modelo funcionará bien con nuevos estudiantes que no fueron parte del conjunto de entrenamiento.
5. **Generación del modelo entrenado:** una vez completado el entrenamiento, se guarda el modelo en un archivo. Este modelo entrenado es lo que posteriormente utilizará el sistema de recomendación para hacer predicciones cuando nuevos estudiantes realicen la prueba de evaluación.

Configuración del entrenamiento:

El modelo se entrena con la siguiente configuración de hiperparámetros:

- **Número de iteraciones:** 400 iteraciones de mejora.
- **Profundidad máxima:** la estructura del árbol de decisión tiene una profundidad máxima de 4 niveles, lo que asegura que el modelo sea interpretable y no memorice datos irrelevantes.
- **Tasa de aprendizaje:** 0.06, lo que significa que el modelo realiza ajustes relativamente pequeños en cada iteración, permitiendo un aprendizaje más estable.
- **Regularización:** se aplican técnicas de regularización para penalizar modelos demasiado complejos, evitando el sobreajuste.

Métrica de evaluación:

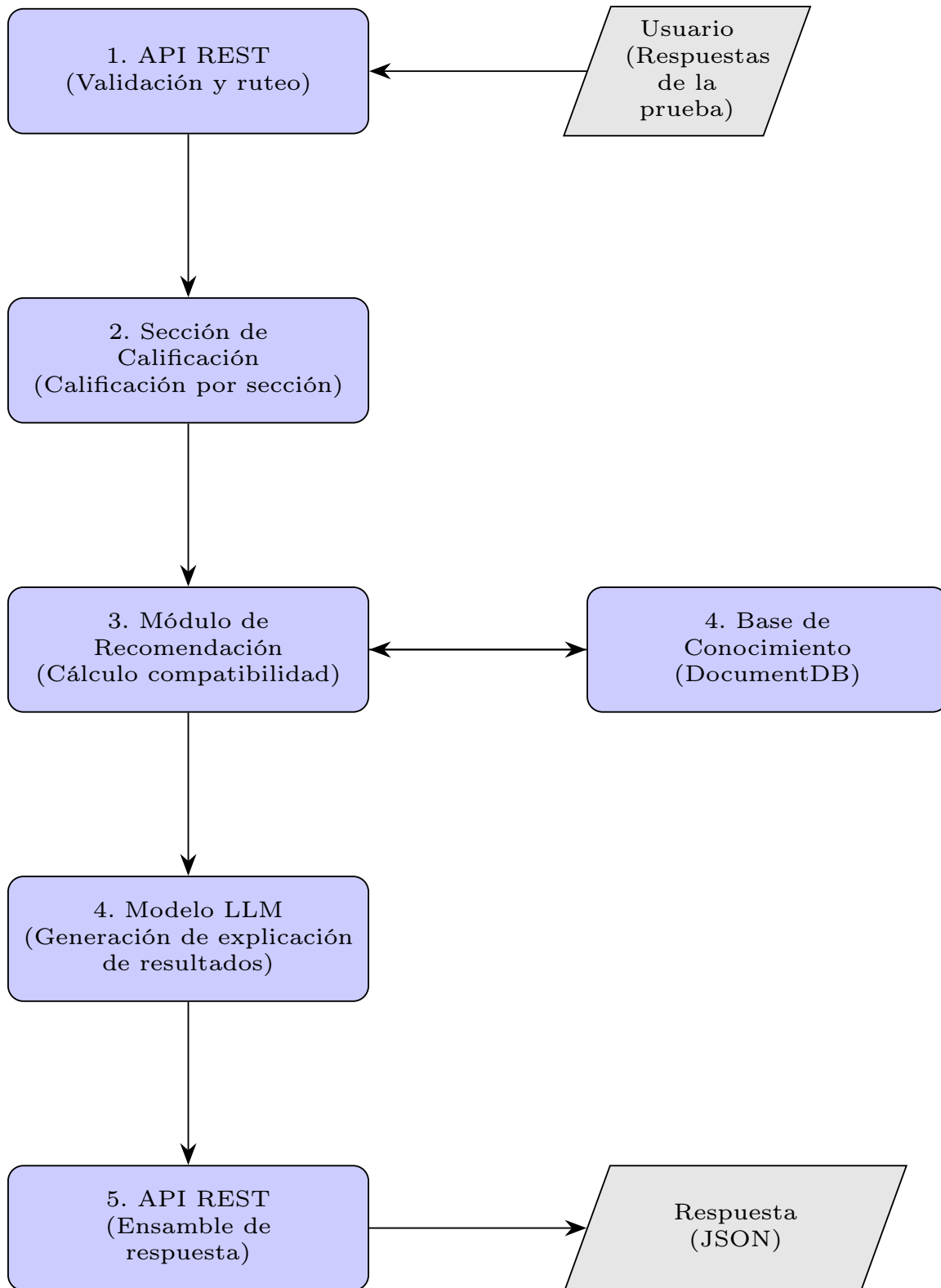
Durante el entrenamiento, se utiliza la métrica **NDCG@5 (Normalized Discounted Cumulative Gain at 5)** (sec. 5.16.7). Esta métrica mide qué tan bien el modelo ordena las mejores carreras recomendadas para un estudiante, con énfasis en las primeras posiciones [101]. Específicamente, evalúa si las cinco primeras recomendaciones generadas por el modelo son las más apropiadas para ese estudiante.

Resultado del entrenamiento:

El resultado es un modelo entrenado que es capaz de hacer recomendaciones de carreras universitarias, sin embargo, por la falta de datos suficientes, el modelo no es óptimo para realizar recomendaciones válidas. Aunque no son recomendaciones que esten basadas en datos reales, si demuestra la viabilidad del modelo de recomendación.

Funcionamiento del modelo de recomendación

Figura 25. Flujo de ejecución del modelo de recomendación



Justificar las recomendaciones de carreras universitarias

Parte del éxito de los resultados obtenidos yace en la capacidad del modelo para justificar y explicar las recomendaciones de carreras universitarias a través de métricas interpretables. Esto se convierte en un aspecto crítico al entender que algunos modelos de machine learning funcionan como cajas negras (sec. 5.18.3), proporcionando predicciones sin justificación visible del proceso de razonamiento.

Sin embargo, XGBoostRanker, al ser un modelo basado en árboles de decisión con gradient boosting, posee la capacidad de extraer y comunicar la **importancia de características** (sec. 5.2.4) que contribuyeron a cada clasificación. De esta manera, el sistema no solo proporciona una recomendación, sino que también puede explicar cuáles atributos del perfil del usuario fueron decisivos para posicionar una carrera en un puesto específico del ranking de recomendaciones. Esta interpretabilidad inherente refuerza la confianza del usuario en el sistema y alinea el modelo con prácticas de IA responsable, al ofrecer explicaciones claras y accesibles de las recomendaciones del modelo 5.18.3. Para ello, se utilizará un LLM, que reciba dichas características, y genera una explicación de las recomendaciones de carreras universitarias basado en el perfil del usuario.

Aquí se cumple parcialmente el **objetivo específico 6 (sec. 3.2)**, de desarrollar e implementar el prototipo de un sistema de recomendación capaz de generar sugerencias de carreras universitarias y contenido académico basado en el perfil del usuario.

Algoritmo de recomendación de contenido

Una vez que el modelo de machine learning ha recomendado las carreras universitarias compatibles con el perfil del estudiante, se le agrega una serie de etiquetas al perfil del estudiante para que el sistema entienda sus preferencias. El siguiente paso es personalizar el contenido que se presenta al usuario, para ello se utiliza un algoritmo de recomendación de contenido que es responsable de seleccionar y ordenar las tarjetas de información (testimonios, detalles de carreras, encuestas, preguntas hipotéticas) de manera que sean relevantes para el estudiante.

Funcionamiento del algoritmo de recomendación de contenido

Pasos del algoritmo:

1. **Identificación del perfil del usuario:** el algoritmo comienza recuperando el perfil del estudiante del sistema, específicamente los **tags (etiquetas)** asociados a su perfil. Estos tags representan los intereses, habilidades y características principales del estudiante extraídas de la prueba de evaluación (por ejemplo: “liderazgo”, “creatividad”, etc.).
2. **Filtrado por relevancia:** el sistema busca en la base de datos todas las tarjetas de contenido disponibles y las filtra manteniendo solo aquellas que contienen al menos una etiqueta que coincida con las del perfil del estudiante. De esta manera, se descartan automáticamente las tarjetas que no son relevantes para este estudiante en particular.

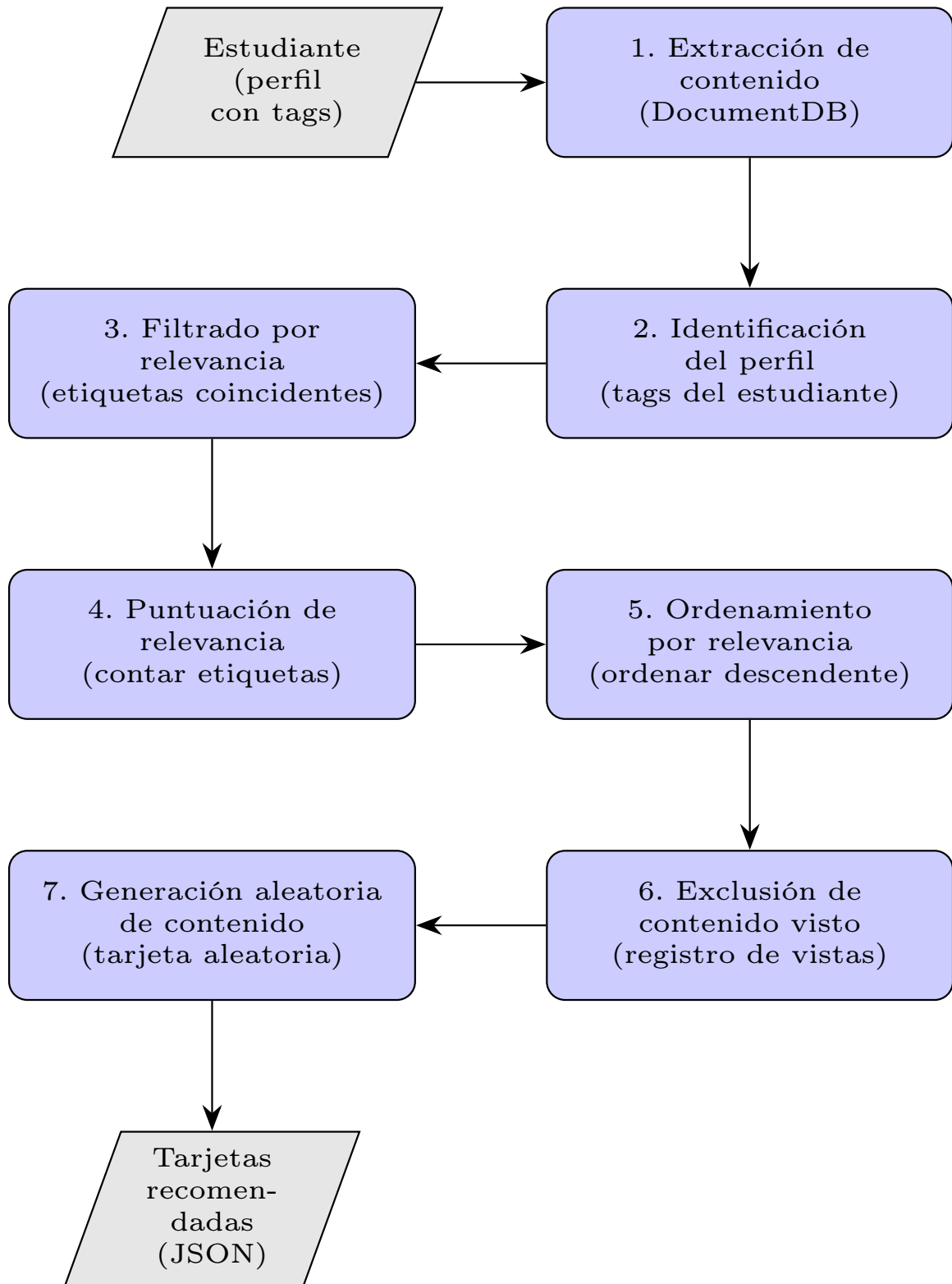
3. **Puntuación de relevancia:** para cada tarjeta que pasó el filtro, el algoritmo calcula una **puntuación de relevancia** contando cuántas etiquetas de la tarjeta coinciden con las etiquetas del estudiante.

Por ejemplo:

- Si una tarjeta sobre “Ingeniería Civil” tiene las etiquetas “análisis”, “precisión” y “trabajo en equipo”, y el estudiante tiene estas 3 etiquetas, la puntuación es 3.
 - Si otra tarjeta sobre “Diseño Gráfico” tiene las etiquetas “creatividad” y “precisión”, y el estudiante tiene ambas, la puntuación es 2.
4. **Ordenamiento por relevancia:** las tarjetas se ordenan de mayor a menor puntuación de relevancia, asegurando que las tarjetas más relevantes aparezcan primero. Esto garantiza que el estudiante ve primero el contenido que más le interesa.
 5. **Exclusión de contenido visto:** el algoritmo mantiene un registro de las tarjetas que el estudiante ya ha visto. En cada solicitud de nuevo contenido, se excluyen automáticamente las tarjetas ya presentadas, garantizando que el estudiante siempre reciba contenido nuevo y fresco.
 6. **Generación aleatoria de contenido dinámico:** al final de cada sesión de recomendación, el algoritmo genera de forma aleatoria una tarjeta adicional con una carrera aleatoria. Esto ayuda al estudiante a explorar otros campos fuera de los recomendados.

Flujo del algoritmo de recomendación de contenido

Figura 26. Flujo del algoritmo de recomendación de contenido



Beneficios del algoritmo:

- **Personalización:** cada estudiante recibe contenido específicamente adaptado a sus características, no contenido genérico.
- **Relevancia:** las tarjetas se ordenan según qué tan bien coinciden con el perfil del estudiante, mostrando primero lo más importante.
- **Diversidad:** la mezcla de diferentes tipos de contenido mantiene la experiencia variada e interesante.
- **Prevención de repetición:** el estudiante nunca ve la misma tarjeta dos veces en la misma sesión.
- **Dinamismo:** la generación aleatoria de contenido asegura que múltiples sesiones siempre presentan contenido nuevo.

Aquí se cumple completamente el **objetivo específico 6 (sec. 3.2)**, al implementar el prototipo de el sistema de recomendación, capaz de generar sugerencias de carreras universitarias y contenido académico basado en el perfil del usuario.

7.4.3. Pruebas funcionales

Con el objetivo de verificar el correcto funcionamiento del flujo de ejecución del modelo, se propondrá la realización de las siguientes pruebas: **(i) ejecución completa del flujo inicial**, **(ii) interoperabilidad con otros módulos** y **(iii) tasa de éxito en las ejecuciones**. Estas evaluaciones permitirán verificar que el flujo de ejecución del modelo se realice sin errores. Dado que el modelo se implementará con datos simulados (*dummy data*), las pruebas no se enfocarán en la coherencia o precisión de las recomendaciones, sino en la comprobación de la funcionalidad y estabilidad general del módulo dentro del sistema.

Diseño de la prueba (marco de prueba del objetivo 7 - 3.2)

Para la validación se propondrá un conjunto de **20 ejecuciones controladas** del módulo de recomendación, utilizando distintos perfiles simulados de usuario con combinaciones de variables de aptitudes e intereses. Cada ejecución seguirá el flujo mostrado en la Figura 25.

1. **Inicialización del perfil:** se enviará un objeto JSON con valores simulados de las 47 variables psicométricas al endpoint `/initial-test/calculate`.
2. **Calificación de los resultados:** se verificarán las respuestas enviadas y se creará un JSON con los resultados por sección para darlos como input al modelo.
3. **Recomendación de carreras:** el sistema calculará las carreras recomendadas utilizando el modelo de machine learning y las calificaciones por sección.

4. **Generación de resultados:** se verificará que el modelo devuelva una lista ordenada de carreras recomendadas.
5. **Recepción e integración:** la API deberá retornar la respuesta correctamente al cliente, mostrando los resultados en formato JSON.

Criterios de éxito del flujo

Para validar el correcto funcionamiento del módulo de recomendación, se establece una meta de **tasa de éxito del 95 %** en las ejecuciones de prueba. Una ejecución se considera exitosa si cumple con todos los criterios definidos en el Cuadro 16. Este umbral fue determinado en función del **objetivo específico 7 (sec. 3.2)**.

Cuadro 16. Criterios de aceptación para el módulo de recomendación

Métrica	Criterio de éxito	Criterio de fallo
Estado de la ejecución	La API retorna un estado HTTP 200 OK.	La API retorna un estado de error (ej. 500).
Tiempo de respuesta	Menor a 25 segundos.	Mayor o igual a 25 segundos.
Contenido	La respuesta contiene 5 carreras recomendadas.	Contiene menos de 5 carreras o ninguna.
	El objeto JSON con los resultados del usuario está presente.	El objeto JSON de resultados está ausente.

La tasa de éxito general del módulo se calculará mediante la siguiente ecuación:

$$\text{Tasa de éxito} = \frac{\text{Nº de ejecuciones exitosas}}{\text{Nº de ejecuciones totales}} \times 100 \quad (5)$$

El porcentaje de éxito general del módulo se calculará mediante la siguiente ecuación:

$$\text{Éxito del flujo} = \frac{\text{Ejecuciones exitosas}}{\text{Ejecuciones totales}} \times 100 \quad (6)$$

Criterio de aceptación

Se establecerá como criterio de aceptación que el módulo cumpla con los requerimientos funcionales si el porcentaje de éxito del flujo de recomendación es igual o superior al **95 %**.

Para la ejecución de la prueba, se utilizará la herramienta de **Postman (sec. 5.17.7)**, utilizando la función de *Collection Runner* (sec. 5.17.7), con 20 ejecuciones seguidas en un intervalo de 35 segundos entre cada ejecución. Se utilizaron scripts (sec. 5.1.3) para la definición automáticas de las pruebas que deberá cumplir cada uno de los requests. Los scripts utilizados se encuentran en el anexo 36.

7.4.4. Resultados de la prueba

Cuadro 17. Resultados de la ejecución de la prueba del objetivo de la recomendación

User	Estado de la ejecución	Tiempo de respuesta	Carreras recomendadas	Resultados del test
User 1	✓	✓	✓	✓
User 2	✓	✓	✓	✓
User 3	✓	✓	✓	✓
User 4	✓	✓	✓	✓
User 5	✓	✓	✓	✓
User 6	✓	✓	✓	✓
User 7	✓	✓	✓	✓
User 8	✓	✓	✓	✓
User 9	✓	✓	✓	✓
User 10	✓	✓	✓	✓
User 11	✓	✓	✓	✓
User 12	✓	✓	✓	✓
User 13	✓	✓	✓	✓
User 14	✓	✓	✓	✓
User 15	✓	✓	✓	✓
User 16	✓	✓	✓	✓
User 17	✓	✓	✓	✓
User 18	✓	✓	✓	✓
User 19	✓	✓	✓	✓
User 20	✓	✓	✓	✓

Nota. Elaboración propia.

El Cuadro 17 muestra los resultados de la ejecución de la prueba para cada uno de los 20 usuarios falsos. La captura de los resultados del test para cada uno de los usuarios se encuentran en las figuras 40, 41, 42 y 43.

Cuadro 18. Resultados de la prueba del objetivo de la recomendación

Prueba	Porcentaje de éxito
Estado de la Ejecución	100 %
Tiempo de Respuesta	100 %
Carreras Recomendadas	100 %
Resultados del Test	100 %

Nota. Elaboración propia.

Con base en el Cuadro 18, el sistema alcanzó una tasa de éxito operativa de **100 %** (20/20 ejecuciones) en la primera interacción simulada, cumpliendo consistentemente los cuatro criterios: (i) estado HTTP 200, (ii) tiempo de respuesta < 25 s, (iii) entrega de 5 carreras recomendadas y (iv) retorno de los resultados del test en un 100 %

7.5. Módulo de comunicación (obj. específicos 8, 9 y 10 - 3.2)

El módulo de comunicación es el encargado de la comunicación entre los diferentes módulos del sistema y el usuario. Para ello, se propone la implementación de una **API RESTful** creada con **TypeScript** y el framework **Hono** que permita orquestar todas las interacciones necesarias para cumplir con los pedidos del usuario.

7.5.1. Justificación de elecciones técnicas

Justificación de la elección de API RESTful

- **Simplicidad:** menor tiempo de desarrollo tomado para comenzar a utilizarla, debido a su facilidad de implementación y mantenimiento.
- **Curva de aprendizaje:** la curva de aprendizaje es baja, lo hace excelente para proyectos pequeños y medianos; que no requieran de una implementación compleja.
- **Familiaridad:** experiencia previa del investigador en la creación y mantenimiento de API RESTful, en proyectos anteriores.
- **Costo:** menor cantidad de recursos necesarios para el funcionamiento de la API, lo que reduce los costos de mantenimiento y actualización.

En consecuencia, se seleccionó API RESTful por su simplicidad, curva de aprendizaje baja, familiaridad y costo, lo que la hace una excelente opción para el módulo de comunicación del sistema propuesto.

Justificación de la elección de TypeScript

- **Curva de aprendizaje:** la curva de aprendizaje es baja, lo hace excelente para proyectos pequeños y medianos; que no requieran de una implementación compleja.
- **Tipado estático:** mejora la mantenibilidad y disminuye la posibilidad de errores. Asegura la calidad del código y mejora la legibilidad.
- **Ecosistema:** ecosistema Node con mucha madurez en el desarrollo de APIs y con una gran cantidad de documentación y recursos para el desarrollo de APIs.
- **Rendimiento:** compila a JavaScript sin coste de rendimiento, mejorando la experiencia de desarrollo, sin sacrificar el rendimiento.
- **Familiaridad:** experiencia previa del investigador en el lenguaje, lo que facilita el desarrollo de la API.

En consecuencia, se seleccionó TypeScript por su curva de aprendizaje baja, tipado estático, ecosistema Node con gran madurez de desarrollo y con una gran cantidad de documentación y recursos. Esto ofrece una buena experiencia de desarrollo, sin sacrificar rendimiento, agregado a la familiaridad que el investigador posee con el lenguaje, lo que facilita el desarrollo de la API.

Justificación de la elección de Hono

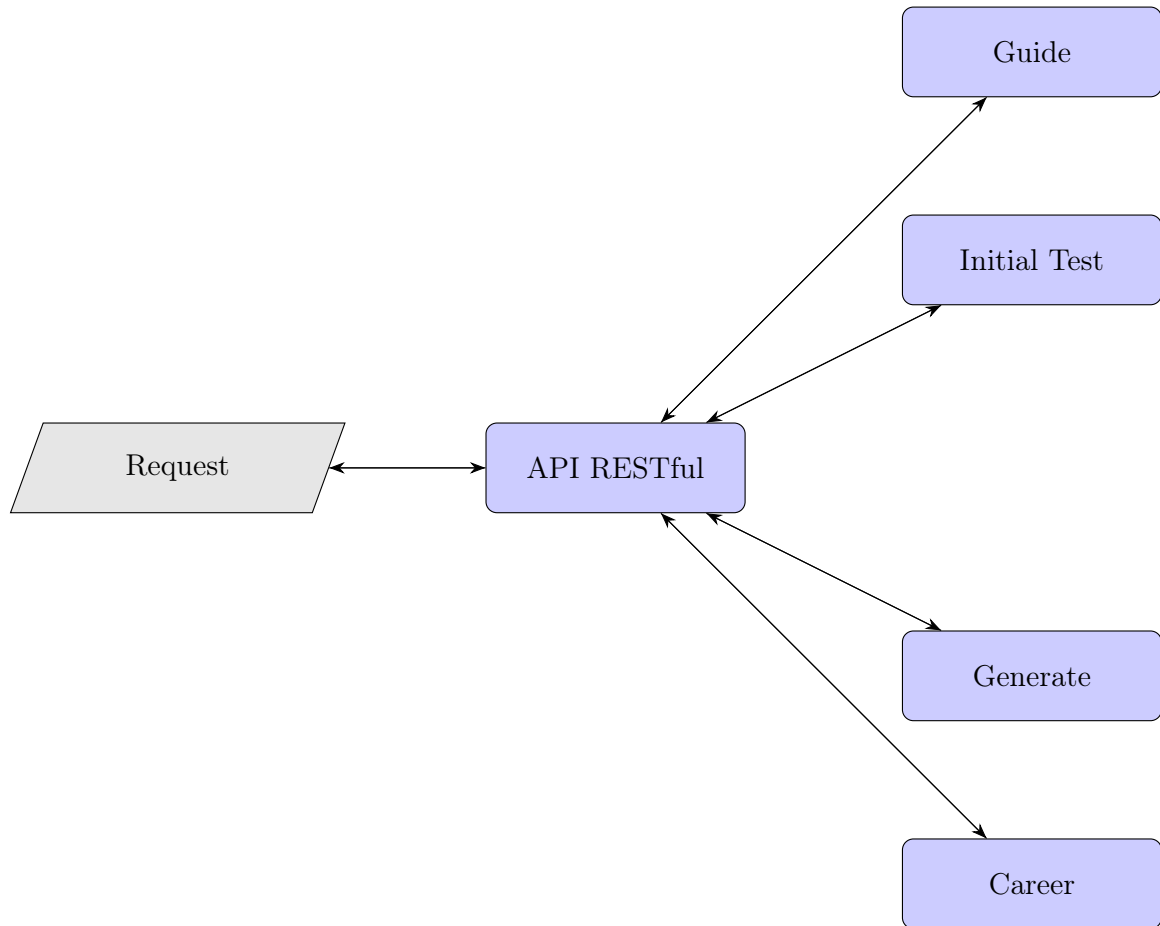
- **Rendimiento:** es uno de los frameworks más rápidos, maneja miles de peticiones con facilidad.
- **Curva de aprendizaje:** es fácil comenzar a utilizarlo, con la curva más baja de todos los frameworks comparados.
- **Poder de cómputo necesario:** su diseño ligero consume pocos recursos, sin sacrificar rendimiento.
- **Familiaridad:** experiencia previa del investigador en el framework, lo que facilita el desarrollo de la API.
- **Buena documentación:** tiene una gran cantidad de documentación y recursos para el desarrollo de APIs.
- **Compatibilidad con librería de Node.js:** tiene una gran cantidad de integraciones con la vasta librería de Node.js, facilitando el desarrollo de la API.

En consecuencia, se seleccionó Hono por su rendimiento alto, curva de aprendizaje baja, poder de cómputo necesario bajo, familiaridad, buena documentación y compatibilidad con librería de Node.js, lo que lo hace ideal para proyectos pequeños y medianos.

7.5.2. Desarrollo

7.5.3. Arquitectura de la API

Figura 27. Diagrama de la arquitectura de la API de comunicación



7.5.4. Endpoints de la API

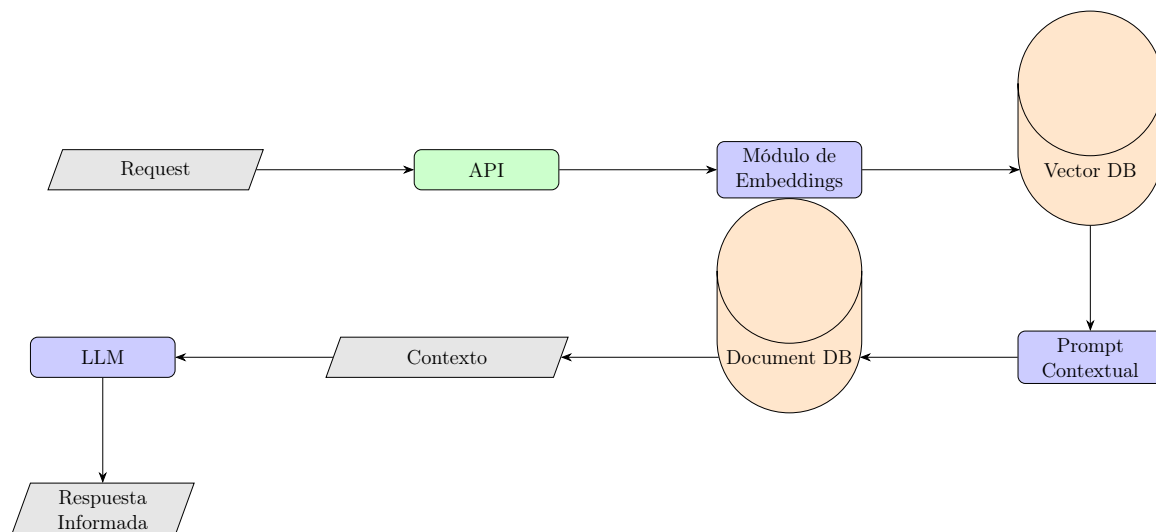
- **/initial-test:** endpoints relacionados con la prueba inicial de evaluación del usuario. Ayuda a satisfacer el **objetivo específico 5 (sec. 3.2)**, ya que es el que provee y genera el perfil del usuario con las respuestas de la prueba inicial de evaluación.
 - **GET /initial-test/items:** se encarga de obtener los ítems de la prueba inicial de evaluación del usuario.
 - **POST /initial-test/calculate:** se encarga de obtener las recomendaciones de carreras universitarias utilizando los resultados de la prueba inicial de evaluación del usuario.
- **/guide:** endpoints relacionados con la guía de preguntas del usuario. Ayuda a satisfacer el **objetivo específico 3 (sec. 3.2)**, ya que es el que provee las rutas necesarias para

el intercambio de preguntas y respuestas con el usuario.

- **POST /guide/ask_without_context:** se encarga de responder al usuario sin contexto.
 - **POST /guide/ask:** se encarga de responder al usuario con contexto de la base de datos de información académica.
- **/generate:** se encarga de generar contenido temático basado en las características del usuario. Ayuda a satisfacer el **objetivo específico 6 (sec. 3.2)**, ya que es el que provee la funcionalidad de generación de contenido temático del usuario.
- **POST /generate/career-combination/:careerId:** se encarga de obtener las preguntas de la generación de contenido temático del usuario.
 - **POST /generate/mini-inform:** se encarga de obtener las respuestas de la generación de contenido temático del usuario.
 - **POST /generate/card:** se encarga de obtener las respuestas de la generación de contenido temático del usuario.
- **/career:** se encarga de recomendar carreras universitarias basadas en las características del usuario. Ayuda a satisfacer el **objetivo específico 2 (sec. 3.2)**, ya que es el que ayude a proveer carreras e información relevante para el usuario utilizando similitudes semánticas.
- **POST /career/similar/:careerId:** se encarga de obtener las preguntas de la recomendación de carreras universitarias del usuario.

7.5.5. Flujo de comunicación entre modelos y bases de datos

Figura 28. Diagrama de flujo RAG - Pregunta del usuario hasta respuesta del LLM



La API actúa como orquestador del flujo de datos. Cuando el usuario envía una consulta, esta se envía al API, que comunica el texto con el **módulo de embeddings**. Una vez calculado el vector, la API consulta la base vectorial para obtener los ids de los documentos más relevantes, luego se recuperan los documentos dentro de la base de datos de información académica. Con los documentos se construye el **prompt contextualizado** y se envía al **LLM** para la generación de la respuesta final.

Con esto se cumple el **objetivo específico 8 (sec. 3.2)**, al implementar un sistema orquestador que maneje la comunicación entre los módulos del sistema.

7.5.6. Pruebas funcionales

Si bien parte de las pruebas experimentales siguientes fueron ejecutadas dentro del contexto de la aplicación móvil desarrollada por otro compañero, en el presente trabajo se analizan exclusivamente aquellos resultados que guardan relación directa con los módulos de consulta y procesamiento de información basados en modelos de inteligencia artificial. Estas pruebas, aunque realizadas en un entorno integrado, permiten evaluar de manera indirecta el desempeño y la pertinencia de las respuestas generadas por el sistema de orientación vocacional, así como su impacto en la experiencia de uso del estudiante. Los resultados no vinculados al funcionamiento del modelo —como los relativos a la interfaz gráfica, navegación o aspectos puramente visuales— se consideran fuera del alcance de esta investigación y se documentan únicamente a modo de referencia contextual.

Diseño de la prueba

Se ejecutó una prueba de recorrido guiado por tareas, registrando para cada interacción: **código de éxito** (E2/E1/E0), **tiempo de ejecución**, **errores** y **observaciones**. Los códigos operativos utilizados fueron:

- **E2**: éxito completo. El participante completó la tarea sin ayuda, de manera fluida y correcta.
- **E1**: éxito parcial. Completó la tarea con dificultad, dudas visibles o correcciones mínimas.
- **E0**: falla total. No completó la tarea o se rindió; necesitó ayuda directa.

Escenarios evaluados

Las tareas se agruparon por módulo del sistema:

- **Autenticación**: crear cuenta con correo; cerrar e iniciar sesión.
- **Quiz**: completar todas las páginas del cuestionario; leer resultados.
- **Exploración**: deslizar tarjetas; abrir/cerrar una tarjeta; revisar Top 5 carreras.

- **Match:** abrir una carrera del Top.
- **Carreras:** explorar pensum interactivo; utilizar “¿y si combino con...?”; filtrar por facultad.
- **Chat:** enviar una pregunta y recibir respuesta; crear nueva conversación; borrar conversación.
- **Perfil:** expandir/editar perfil; retomar quiz; borrar cuenta.
- **General:** navegación libre; comentario de experiencia general.

Para más información, se puede consultar sobre las tareas se puede visitar el anexo 44 y para saber que aspectos se midieron se puede visitar 45. Esto cumple con el **objetivo específico 9**, al crear las tareas para la prueba del funcionamiento del sistema.

Métricas recogidas y criterios

- **Errores:** conteo cualitativo de incidencias visibles durante la interacción.
- **Satisfacción:** escala Likert de 1–5 por módulo (UX/HCI: usabilidad, estética, claridad, eficiencia, coherencia).

7.5.7. Resultados de la prueba

A continuación se presentan los resultados de la prueba realizada a 6 usuarios, las métricas se encuentran en el anexo 11.1.5. Como nuestro objetivo es evaluar las respuestas de nuestro sistema y la información que se le proporciona al usuario, estaremos evaluando el éxito y satisfacción de módulos de Exploración, Quiz, Match, Carreras, Chat y Perfil.

Cuadro 19. Resultados promediados de la prueba funcional por módulo: códigos E2/E1/E0 y satisfacción (escala 1–5)

Módulo	E2	E1	E0	Satisfacción
Exploración	100.00 %	0.00	0.00	86.6 %
Quiz	100.00 %	0.00	0.00	100.00 %
Match	100.00 %	0.00	0.00	93.4 %
Carreras	94.44 %	5.56 %	0.00	90.0 %
Chat	94.44 %	5.56 %	0.00	93.4 %
Perfil	100.00 %	0.00	0.00	100.00 %
Total	98.14 %	0.019 %	0.00	93.90 %

En el Cuadro 19 se puede observar una recopilación de todos los resultados, relevantes para nosotros. Aquí se puede ver que en la **gran mayoría de secciones se tuvo éxito** en la ejecución de la tarea, solo se **experimentaron fallos en las secciones de Carreras y Chat**; obteniendo un éxito en tareas del **98.14 % a nivel general**.

En la parte de la satisfacción, se puede ver que se obtuvo una **satisfacción por encima del 85 % en todas las secciones**, siendo Exploración la más baja con 86 %. Sin embargo, se nota que solo la sección de **Quiz y Perfil** fueron las únicas que alcanzaron un **nivel de satisfacción mayor al del 95 %**, lo que explica que **nivel de satisfacción general sea del 93.9 %**.

8.1. Encuesta inicial

1. Los hallazgos de la encuesta son más representativos de la realidad urbana y ladina/mestiza, y podrían no capturar de manera adecuada la perspectiva de estudiantes de pueblos originarios u otras comunidades con dinámicas vocacionales distintas.
2. La composición demográfica de la población encuestada muestra que la mayoría de los estudiantes viven en la Ciudad Capital de Guatemala, lo cuál permite un mayor sesgo en los resultados de la encuesta hacia las preferencias y expectativas de residentes de la Ciudad Capital de Guatemala.
3. Existe una cantidad considerable de estudiantes que todavía no han logrado explorar todas las opciones que tienen disponibles.
4. Los estudiantes de bachillerato no pueden depender únicamente de su centro educativo para informarse sobre sus opciones de carrera.
5. Persiste un alto grado de incertidumbre y falta de seguridad en el proceso de elección por parte de los estudiantes.
6. Aunque una gran cantidad de jóvenes han tenido un buen acercamiento y una actitud de búsqueda activa sobre sus opciones académicas, existe un porcentaje considerable de estudiantes que no se han sentido motivados para prepararse para tomar esta decisión.
7. El uso emergente de la inteligencia artificial, refleja una apertura de los jóvenes a incorporar nuevas tecnologías en sus procesos de búsqueda vocacional.
8. La mayoría percibe que aún les falta información en algún aspecto, que podría significar un cambio en la elección de carrera que puedan tomar.
9. Aunque la mayoría de los jóvenes ha tenido algún tipo de acercamiento a la orientación vocacional, todavía existe una proporción relevante de estudiantes sin apoyo suficiente.

10. Los estudiantes valoran la personalización, la confiabilidad y la interactividad más que la mera disponibilidad horaria.
11. Existe una alta aceptación potencial hacia el uso de herramientas digitales de orientación vocacional.
12. El reto principal no radica en la falta de acceso tecnológico, sino en la ausencia de orientación inicial clara y estructurada.

8.2. Módulo de información (obj. específicos 1 y 2 - 3.2)

1. Los resultados de la prueba de integridad estructural (sec. 7.2.6) demuestran que la base de datos no posee documentos incompletos, lo que cumple con el criterio de aceptación **del objetivo específico 1 (sec. 3.2)**, probando que el 100 % de las carreras en la base de datos cumplen con la estructura esperada.
2. Los resultados de la prueba de precisión de consulta (sec. 7.2.6) demuestran que los datos de cada carrera dentro de la base de datos es correcta con la información registrada en el sitio oficial de la Universidad del Valle de Guatemala hasta la fecha cuando se está redactando este trabajo (30 de octubre de 2025). Esto cumple con el criterio de aceptación **del objetivo específico 2 (sec. 3.2)** de alcanzar una precisión mínima del **95 %** de precisión de consulta de la base de datos, asegurando que el sistema cuenta con la información correcta para proveer al sistema. Las consultas realizadas y sus resultados se encuentran en el anexo 11.1.1.
3. La base de conocimiento provee un marco sólido y escalable para la integración de actualización e integración de nueva información.
4. La validez depende de que la fuente oficial no cambie; riesgo de desactualización de la información.
5. Tamaño de muestra de preguntas podría ser insuficiente para cubrir casos límite.
6. La base de datos muestra insuficiencia de información de los cursos de las carreras, por ser datos no públicos administrados por la UVG.

8.3. Módulo de guianza vocacional (obj. específicos 3 y 4 - 3.2)

1. La implementación del sistema de guianza vocacional fue exitosa, cumpliendo con el **objetivo específico 3 (sec. 3.2)**, de implementar un sistema de guianza vocacional basado en un Modelo de Lenguaje Extenso (LLM) que utilice la base de conocimiento construida para responder preguntas sobre carreras universitarias.
2. El sistema de guianza vocacional cumple con el **objetivo específico 4 (sec. 3.2)**, con un desempeño alto y consistente (promedio 4.89/5; fidelidad 97.78 %), supera con holgura el umbral del 85 % establecido en el objetivo específico y el criterio de aceptación.

Las respuestas son factualmente correctas (4.92), relevantes (4.85) y claras/completas (4.9).

3. El trabajo permitió entender y aplicar la utilización de LLMs, a través de APIs, para la generación de respuestas y la creación de embeddings.
4. La implementación del sistema de guianza vocacional permitió entender y aplicar enfoques y tecnologías implementadas actualmente en el mercado, como el enfoque Retrieval-Augmented Generation (RAG), para la recuperación de información relevante. Entendiendo los beneficios y limitaciones de este enfoque.
5. La creación del enfoque RAG permitió entender y aplicar la utilización de Vector DBs, a través del embeddings de preguntas y la búsqueda vectorial, para la recuperación de información relevante.
6. El proceso de implementación del modelo de recomendación, me llevo a aprender como agregar contexto e instrucciones para comunicarme con los los LLMs, de manera que pueda obtener respuestas más informadas y al punto.
7. Cuando se requiere de contexto para la generación de respuestas, comprendí que si se le proporciona demasiada información al modelo, puede llegar a sobrecargarse y solo recuperar información irrelevante, dejando detalles importantes sin responder. Por lo que se debe de intentar minimizar la cantidad de información que se le proporciona al modelo, solo proporcionando la información relevante para la pregunta.
8. Durante la implementación del sistema, se aprendió las ventajas y desventajas de la desnormalización de datos, con la necesidad de rápida recuperación de los datos relevantes.
9. En la pruebas realizadas, se pudo ver una falta en la recuperación de información relevante del historial de consultas, necesitando que se le indicara nuevamente las carreras para contestar las preguntas.

8.4. Módulo de recomendación (obj. específicos 5, 6 y 7 - 3.2)

1. El sistema de recomendación cumple con el **objetivo específico 6 (sec. 3.2)**, de desarrollar e implementar el prototipo de un sistema de recomendación capaz de generar sugerencias de carreras universitarias y contenido académico basado en el perfil del usuario.
2. El sistema de recomendación cumple el **objetivo específico 7 (sec. 3.2)**, de obtener una tasa de éxito en el flujo de recomendación inicial superior al 95 %. Lo que logra demostrar la **viabilidad del prototipo** para ejecutar el flujo extremo a extremo y generar recomendaciones de carreras universitarias y contenido académico basado en el perfil del usuario.
3. La prueba valida estabilidad y orquestación del flujo inicial; no evalúa la calidad de las recomendaciones al usar datos simulados.

4. La prueba valida la correcta interacción entre los diferentes módulos del sistema, demostrando la interoperabilidad entre ellos.
5. Se entendió que el éxito de una persona en un campo no solo depende de su capacidad cognitiva y las habilidades adquiridas, sino de aspectos como la personalidad, la perseverancia y los intereses vocacionales.
6. Se entendió que el éxito dentro de un campo no solo debería basarse en la capacidad de laborar, sino en la satisfacción que la persona encuentra al realizar dicho trabajo.
7. En la implementación del sistema de recomendación, se aprendió a ejecutar modelos de Machine Learning, en lenguaje Python, desde una API que utiliza un lenguaje de programación distinto.
8. Se aprendió la estructura del dataset necesario para el entrenamiento del modelo de Machine Learning XGBoostRanker, para la clasificación de carreras universitarias.
9. Se aprendió a interpretar el score del ranking y las razones por cada carrera universitaria.
10. Las principal limitación fue la falta de un conjunto de datos real para el entrenamiento del modelo de recomendación, lo cual impidió evaluar métricas de precisión o recall.
11. Se entendió que el diseño actual del modelo facilita mucho su implementación, bajando la complejidad necesaria para su entrenamiento y ejecución. Sin embargo, se entiende que a menor complejidad del modelo, mayor es la cantidad y complejidad que debe de tener el dataset para poder entrenar el modelo de manera efectiva.

8.5. Módulo de comunicación (obj. específicos 8, 9 y 10 - 3.2)

1. La prueba funcional 7.4.3 valida la correcta orquestación del flujo de recomendación inicial, al cumplir con los criterios de éxito definidos en la Tabla 16, demostrando una orquestación operativa entre los módulos, no solo funcional, sino eficiente al realizar con los criterios de éxito definidos. Cumpliendo con el **objetivo específico 8 (sec. 3.2)** al comprobar su exitosa implementación operativa.
2. Con una satisfacción del 93.9% del sistema general, observado en la cuadro 19, y con un porcentaje de tareas ejecutadas exitosamente del 98.14 %; se cumplen con los requerimientos del **objetivo específico 10 (sec. 3.2)**.
3. El desarrollo modular del sistema demostró que facilita la integración entre los diferentes módulos del sistema, permitiendo una mayor mantenibilidad del sistema, y, a su vez, facilitando el cambio de alguna herramienta o tecnología.
4. El proceso de implementación fortaleció mis habilidades en integración de APIs, manejo de bases de datos no relacionales, implementación de LLMs dentro del análisis dinámico de información y la creación de un sistema de recomendación basado en características.
5. El sistema de comunicación cumple con el objetivo planteado, y permite comunicar los resultados del sistema de recomendación y guianza vocacional al usuario.

8.6. Conclusión final

A través del test inicial se identificó que si existe una necesidad de realizar un mejoramiento del proceso de orientación vocacional para los estudiantes de bachillerato, ya que existe una cantidad considerable de estudiantes que no han logrado explorar todas las opciones que tienen disponibles, que sienten que los medios tradicionales de orientación vocacional no son suficientes, y existe un porcentaje considerable de estudiantes que no se han sentido motivados para prepararse para tomar esta decisión. Mostrando una disponibilidad a utilizar herramientas digitales que les permitan la continua exploración de los diferentes campos profesionales.

El sistema de recomendación y guía vocacional cumple con el objetivo planteado, y permite recomendar carreras universitarias y contenidos basado en las características de los usuarios, utilizando los datos recolectados de una prueba de evaluación, y responder preguntas relacionadas con carreras universitarias, planes de estudio y contenidos temáticos, utilizando información confiable y actualizada proveniente de las fuentes institucionales. Haciendo uso de una arquitectura modular, se logró una mayor mantenibilidad del sistema y facilidad para el cambio de alguna herramienta o tecnología; integrando al mismo tiempo tecnologías y enfoques modernos que ayudan a la escalabilidad y eficiencia del sistema.

9.1. Recomendaciones para el sistema de recomendación

9.1.1. Encuesta inicial

- Para futuros levantamientos de datos, se recomienda realizar una mayor cantidad de encuestas, con el fin de mejorar la comprensión de las necesidades de los estudiantes.
- Con el propósito de mejorar la diversidad de los resultados, sería pertinente ampliar la muestra a contextos rurales y poblaciones indígenas, para tener opiniones más inclusivas y generalizables.

9.1.2. Módulo de información

- La información de los cursos de las carreras universitarias es muy escasa. Se recomienda expandir la información con información como contenido, prerequisites, habilidades a desarrollar, etc. Considerando que la información de los cursos es fundamental y ayudaría al modelo a contestar las preguntas de manera más precisa, al recibir más contexto sobre los temas que trata la carrera.
- Realizar auditorías y monitoreo de la base de datos para asegurar la integridad y la actualización de la información conforme a la fuente oficial de la UVG.
- Ampliar el set de consultas de prueba para cubrir casos límite y casos que no fueron considerados en el desarrollo del proyecto.

9.1.3. Módulo de guianza vocacional

- Mejorar la recuperación de información relevante, implementando un sistema de caching para la información de las preguntas dentro del historial de consultas, para solo

ir agregando información y no recuperar toda la información nuevamente.

- Mejorar la continuidad de la conversación, al tomar en cuenta el historial completo y poder proporcionar respuestas a preguntas que utilicen información utilizada anteriormente o las mismas respuestas dadas por el modelo.
- Incluir en cada dato introducido dentro de la respuesta la fuente utilizada (de la base de conocimiento), en vez de solo incluirlas todas en un campo aparte con la respuesta.
- Reajustar las tags de preferencia del usuario utilizando las consultas realizadas en el chat.
- Añadir la opción al LLM de interpretar si el usuario quiere que se realice alguna acción dentro de la aplicación, como por ejemplo, al preguntar por carreras que incluyen ciertos temas, el LLM podría indicarle la carrera y podría ofrecer la opción de mandarlo hacia la sección donde se muestre la información completa de la carrera.

9.1.4. Módulo de recomendación

- La prueba es de carácter *operativo/funcional*; al usar datos simulados, no evalúa la calidad de las recomendaciones. Como trabajo futuro, se propone validar con datos reales y ampliar pruebas de carga y robustez para el modelo de recomendación.
- Para mejorar diversidad e inclusión del modelo, se recomienda ampliar la muestra de estudiantes a contextos rurales y poblaciones indígenas.
- En los items de la prueba de evaluación, contamos con una cantidad limitada de items por cada sección. Se recomienda **expandir la cantidad de items en cada sección**, para poder obtener resultados más precisos; esto considerando que intentar medir la complejidad de las características propuestas con la cantidad limitada de items, puede ser insuficiente para obtener resultados suficientes.
- Dentro de las características propuestas, se analizaron temas como la personalidad, las habilidades cognitivas, las habilidades no cognitivas y los intereses vocacionales. Sin embargo, se recomienda **expandir dichas características**, con la **inclusión de una sección donde se evalúe la satisfacción con la carrera de estudiantes hasta la fecha**. Así poder determinar, no solo si la persona es compatible con la carrera, sino si la persona se sentirá satisfecha con las actividades y contenidos que se ofrecen en la carrera.
- La inclusión de resultados de personas profesionales que estén laborando dentro del campo profesional de cada carrera podría ayudar a consolidar el perfil de cada carrera, teniendo ya un enfoque más realista del perfil esperado dentro del mercado laboral.

9.1.5. Módulo de comunicación

- Se recomienda poder implementar caching para información que se consulte muy frecuentemente, como por ejemplo, si el auge de una carrera universitaria es muy alto,

se podría implementar caching para la información de la carrera, evitando consultas repetidas a la base de datos y minimizando el tiempo de respuesta del sistema.

CAPÍTULO 10

Bibliografía

- [1] B. E. Cisneros-Bravo, R. M. Rodríguez-Aguilar, Y. E. Niño-Membrillo y A. D. Cuevas-Rasgado, “Falta de orientación vocacional como factor en la deserción universitaria. Caso de estudio: zona Oriente del Estado de México,” *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, vol. 14, nov. de 2023. DOI: 10.23913/ride.v14i27.1715
- [2] P. Albino, *Deserción universitaria: ¿cuáles son las razones y cómo prevenirla?* SYD-LE, 2022. dirección: <https://www.sydle.com/es/blog/desercion-universitaria-639a22f22ff02745fa4eface>
- [3] P. M. C. Á. Humana, *Desmotivación: ¿Qué es? Causas y cómo combatirla*, Área Humana, 2021. dirección: <https://www.areahumana.es/desmotivacion/>
- [4] Mafer, *Eres del 40 por ciento de universitarios que se equivoca al elegir carrera*, EF Blog Mexico, 2020. dirección: <https://www.ef.com.mx/blog/language/4-de-cada-10-universitarios-cambia-de-carrera-en-mexico/>
- [5] uPlanner, *Qué motiva la deserción estudiantil en México*, Motivación de estudiantes en México, 2016. dirección: <https://uplanner.com/es/que-motiva-la-desercion-estudiantil-en-mexico/>
- [6] E. de eLearning Actual, *Informe del Mercado Global de Educación en Línea 2024-2032 - eLearning Actual*, 2024. dirección: <https://elearningactual.com/informe-del-mercado-global-de-educacion-en-linea-2024-2032/>
- [7] S. Chauhan y N. K. Panda, “Foundation: Understanding the Basics,” *Hacking Web Intelligence*, págs. 1-13, ene. de 2015, Recuperado el: 02/08/2025. DOI: 10.1016/B978-0-12-801867-5.00001-X dirección: <https://linkinghub.elsevier.com/retrieve/pii/B978012801867500001X>
- [8] Github, *What is a programming language?* Recuperado el: 02/08/2025, jun. de 2024. dirección: <https://github.com/resources/articles/software-development/what-is-a-programming-language>

- [9] E. Cantos, *Algoritmo / Programación... ¿una forma de lenguaje?* Recuperado el: 03/08/2025, 2025. dirección: <https://rea.ceibal.edu.uy/elp/programacion-una-forma-de-lenguaje/algoritmo.html>
- [10] A. Nivel, *¿Qué es un algoritmo en informática? Guía para principiantes - Alto Nivel*, Recuperado el: 03/08/2025, feb. de 2024. dirección: <https://www.altonivel.com.mx/que-es-un-algoritmo-en-informatica-guia-para-principiantes/>
- [11] M.-W. Dictionary, *ALGORITHM Definition & Meaning - Merriam-Webster*, Recuperado el: 03/08/2025; p. 1. dirección: <https://www.merriam-webster.com/dictionary/algorithm>
- [12] N. C. S. R. Center, *Script — Glossary Entry*, <https://csrc.nist.gov/glossary/term/script>, Último acceso: 2025, 2025.
- [13] Google, *¿Qué es la inteligencia artificial (IA)? Google Cloud / Google Cloud*, Recuperado el: 03/08/2025; p. 3. dirección: <https://cloud.google.com/learn/what-is-artificial-intelligence>
- [14] S. J. Russell y P. Norvig, *Artificial Intelligence: A Modern Approach*. Pearson, 2020, cap. 1, págs. 19-49, Recuperado el: 03/08/2025.
- [15] A. Vaswani et al., “Attention Is All You Need,” en *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. dirección: <https://arxiv.org/abs/1706.03762>
- [16] T. Mikolov, I. Sutskever, K. Chen, G. Corrado y J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” en *Advances in Neural Information Processing Systems (NeurIPS)*, 2013. dirección: <https://proceedings.neurips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
- [17] J. Pennington, R. Socher y C. D. Manning, “GloVe: Global Vectors for Word Representation,” en *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, págs. 1532-1543. DOI: 10.3115/v1/D14-1162 dirección: <https://aclanthology.org/D14-1162/>
- [18] J. Devlin, M. Chang, K. Lee y K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” en *Proceedings of NAACL-HLT*, 2019, págs. 4171-4186. dirección: <https://aclanthology.org/N19-1423.pdf>
- [19] N. Reimers e I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” *arXiv:1908.10084*, 2019. dirección: <https://arxiv.org/abs/1908.10084>
- [20] Y. Bengio, A. Courville y P. Vincent, “Representation Learning: A Review and New Perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, n.º 8, págs. 1798-1828, 2013. DOI: 10.1109/TPAMI.2013.50 dirección: <https://arxiv.org/abs/1206.5538>
- [21] B. I. D. S. Team. “Understanding Feature Importance in Machine Learning.” “Feature importance refers to techniques for determining the degree to which different features, or variables, impact a machine learning model’s predictions.”. dirección: <https://builtin.com/data-science/feature-importance>

- [22] B. E. Team. “What Is Feature Importance in Machine Learning?” “Feature (variable) importance indicates how much each feature contributes to the model prediction.”. dirección: <https://www.baeldung.com/cs/ml-feature-importance>
- [23] S. Mishra, S. Dutta, J. Long y D. Magazzeni, “A Survey on the Robustness of Feature Importance and Counterfactual Explanations,” *arXiv preprint arXiv:2111.00358*, 2021, Survey that discusses feature-importance methods and their robustness. dirección: <https://arxiv.org/abs/2111.00358>
- [24] H. Enterprise, *What is Artificial Intelligence (AI)? | Glossary | HPE LAMERICA*, Recuperado el: 03/08/2025, feb. de 2025. dirección: <https://www.hpe.com/lamerica/es/what-is/artificial-intelligence.html>
- [25] *Machine Learning | Definition, types, and examples | SAP*, SAP, 2017. visitado 4 de ago. de 2025. dirección: <https://www.sap.com/products/artificial-intelligence/what-is-machine-learning.html>
- [26] SAP, *Machine Learning | Definition, types, and examples | SAP*. dirección: <https://www.sap.com/products/artificial-intelligence/what-is-machine-learning.html>
- [27] IBM, *What Is Machine Learning (ML)? | IBM*, sep. de 2021. dirección: <https://www.ibm.com/think/topics/machine-learning>
- [28] Databricks, *What are Machine Learning Models?* 2022. dirección: https://www.databricks.com/glossary/machine-learning-models?utm_source=chatgpt.com
- [29] G. Cloud, *Deep learning vs machine learning vs AI | Google Cloud*. dirección: <https://cloud.google.com/discover/deep-learning-vs-machine-learning>
- [30] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.
- [31] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016. dirección: <https://www.deeplearningbook.org/>
- [32] T. Hastie, R. Tibshirani y J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2.^a ed. New York: Springer, 2009.
- [33] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” en *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 1995, págs. 1137-1145.
- [34] J. Bergstra e Y. Bengio, “Random Search for Hyper-Parameter Optimization,” *Journal of Machine Learning Research*, vol. 13, págs. 281-305, 2012.
- [35] S. Garcia, S. Ramírez-Gallego, J. Luengo, J. M. Benítez y F. Herrera, “A review on data preprocessing methods for big data,” *Big Data Analytics*, vol. 1, n.º 1, págs. 1-22, 2016.
- [36] M. F. Triola, *Estadística*, 13.^a ed. Madrid, España: Pearson Educación, 2018.
- [37] D. S. Moore, G. P. McCabe y B. A. Craig, *Introduction to the Practice of Statistics*, 10.^a ed. New York, USA: W. H. Freeman y Company, 2021.
- [38] Penn State Department of Statistics. “1.2 – Samples & Populations | STAT 200.” Definiciones de población y muestra, y relación con parámetros y estadísticas. Accedido: 2025-10-18. dirección: <https://online.stat.psu.edu/stat200/lesson/1/1.2>

- [39] OpenStax. “Introductory Statistics 2e: 1.1 Definitions of Statistics, Probability, and Key Terms.” Glosario inicial con definiciones de población, muestra, parámetro y estadístico. Accedido: 2025-10-18. dirección: <https://openstax.org/books/introductory-statistics-2e/pages/1-1-definitions-of-statistics-probability-and-key-terms>
- [40] OpenStax. “Introductory Statistics 2e: 9 Introduction.” Describe la inferencia desde muestras hacia poblaciones. Accedido: 2025-10-18. dirección: <https://openstax.org/books/introductory-statistics-2e/pages/9-introduction>
- [41] G. James, D. Witten, T. Hastie y R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.
- [42] T. S. Frontier, “Orchestration vs. Choreography in Distributed Architectures: A Deep Dive,” *Medium*, 2025, “Orchestration provides centralized control ...”. dirección: <https://medium.com/the-software-frontier/orchestration-vs-choreography-in-distributed-architectures-a-deep-dive-7eee9abdd423>
- [43] A. S. Gillis, “What is Orchestration in IT?” *TechTarget*, 2025, “Orchestration ... coordination and management of different processes, systems and tasks.”. dirección: <https://www.techtarget.com/searchitoperations/definition/orchestration>
- [44] N. I. of Standards y T. (NIST), *Glossary: Service-Oriented Architecture (SOA)*, https://csrc.nist.gov/glossary/term/service_oriented_architecture, “A set of principles and methodologies for designing and developing software in the form of interoperable services.”, 2020.
- [45] I. Corporation. “What is Service-Oriented Architecture (SOA)?” “SOA defines a way to make software components reusable and interoperable through service interfaces.”. dirección: <https://www.ibm.com/think/topics/soa>
- [46] I. Corporation. “What is Prompt Engineering?” Prompt engineering is the process of writing and refining input prompts to guide generative AI systems toward specific, high-quality outputs. dirección: <https://www.ibm.com/think/topics/prompt-engineering>
- [47] I. Amazon Web Services. “What is Prompt Engineering?” Prompt engineering involves designing and optimizing text inputs that guide generative AI models to produce useful and relevant responses. dirección: <https://aws.amazon.com/what-is/prompt-engineering/>
- [48] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi y G. Neubig, “Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing,” *arXiv preprint arXiv:2302.11382*, 2023, This survey provides a comprehensive overview of prompt-based learning and engineering strategies for large language models. dirección: <https://arxiv.org/abs/2302.11382>
- [49] T. E. Team. “Prompt Engineering Definition.” Prompt engineering is an AI engineering technique that refines the interaction between humans and large language models. dirección: <https://www.techtarget.com/searchenterpriseai/definition/prompt-engineering>
- [50] American Psychological Association. “item analysis.” APA Dictionary of Psychology. Procedimientos para evaluar el mérito estadístico de ítems individuales. dirección: <https://dictionary.apa.org/item-analysis>

- [51] J. Muñiz, *Introducción a la psicometría*. Pirámide, 2010.
- [52] C. J. Soto, A. Kronauer y J. K. Liang, “Five-factor model of personality,” vol. 2, págs. 506-510, 2016.
- [53] “Concepts and Coefficients Based on John L. Holland’s Theory of Vocational Choice—Examining the R Package *holland*,” *Psych* 2021, Vol. 3, Pages 728-750, vol. 3, págs. 728-750, 4 nov. de 2021, ISSN: 2624-8611. DOI: 10.3390/PSYCH3040047 dirección: <https://www.mdpi.com/2624-8611/3/4/47/htm%20https://www.mdpi.com/2624-8611/3/4/47>
- [54] S. D. W. Partnership, *RIASEC - San Diego Workforce Partnership*, jul. de 2020. dirección: <https://workforce.org/san-diego-jobs/my-next-move-2/riasec/>
- [55] A. L. Duckworth, C. Peterson, M. D. Matthews y D. R. Kelly, “Grit: Perseverance and Passion for Long-Term Goals,” *Journal of Personality and Social Psychology*, vol. 92, págs. 1087-1101, 6 jun. de 2007, ISSN: 00223514. DOI: 10.1037/0022-3514.92.6.1087
- [56] I. García-Moya, F. Rivera, C. Moreno, B. Lindström y A. Jiménez-Iglesias, “Analysis of the importance of family in the development of sense of coherence during adolescence,” *Scandinavian Journal of Public Health*, vol. 40, págs. 333-339, 4 2012. DOI: 10.1177/1403494812449924
- [57] X. Xing y J. W. Rojewski, “Family Influences on Career Decision-Making Self-Efficacy of Chinese Secondary Vocational Students,” *New Waves Educational Research & Development*, vol. 21, págs. 48-67, 1 2018.
- [58] J. Dietrich y B. Kracke, “Career-specific parental behaviors in adolescents’ development,” *Journal of Vocational Behavior*, vol. 75, n.º 2, págs. 109-119, 2009. DOI: 10.1016/j.jvb.2009.03.005
- [59] O. Kivinen y J. Nurmi, “Socioeconomic status and educational choice,” *Higher Education*, vol. 81, págs. 101-120, 2021. DOI: 10.1007/s10734-020-00601-3
- [60] F. K. Nderitu, B. Mukadi y P. Tarus-Kiptoo, “International Journal of Educational Research and Development www.allinnovationjournal.com Online Peer pressure influence on career decision making self-efficacy among students in public secondary schools in Kiambu county, Kenya,” vol. 6, págs. 16-21, 4 2024, ISSN: 2664-7087. dirección: www.allinnovationjournal.com
- [61] E. J. Porfeli y B. Lee, “Career development during childhood and adolescence,” *New directions for youth development*, vol. 2012, 134 2012, ISSN: 1537-5781. DOI: 10.1002/YD.20011 dirección: <https://pubmed.ncbi.nlm.nih.gov/22826162/>
- [62] J. Libiano, *Las capacidades cognitivas: qué son, tipos, funcionamiento y estimulación*, mar. de 2022. dirección: <https://neuronup.com/estimulacion-y-rehabilitacion-cognitiva/las-capacidades-cognitivas-que-son-tipos-funcionamiento-y-estimulacion/>
- [63] Kumon, *Kumon / Habilidades cognitivas: ¿qué son, tipos y cómo desarrollarlas?* 2022. dirección: <https://www.kumon.com.pe/blog/habilidades-cognitivas/>
- [64] M. Ramos, *Las habilidades no cognitivas y la Universidad - Universidad, sí*, mayo de 2017. dirección: https://www.universidadsi.es/las-habilidades-no-cognitivas-la-universidad/?utm_source=chatgpt.com
- [65] OECD, *PIAAC*. dirección: https://www.oecd.org/en/about/programmes/piaac.html?utm_source=chatgpt.com

- [66] C. Board, PAA. dirección: https://latam.collegeboard.org/paa/?utm_source=chatgpt.com
- [67] V. Gaubau, ¿Qué es una profesión? Jun. de 2018. dirección: <https://www.consejo-colef.es/post/vgaubau-profesion>
- [68] INTERIM-GROUP, *Carrera profesional: estrategias empresariales para desarrollarla - INTERIM GROUP*, jun. de 2024. dirección: <https://interimgrouphr.com/blog/carrera-profesional/>
- [69] Remote, *Todo sobre las trayectorias profesionales*, ago. de 2025. dirección: <https://remote.com/es-es/recursos/glosario-rrhh/definicion-trayectoria-profesional>
- [70] UNIR, *La orientación vocacional: ¿qué es y para qué sirve?* Dic. de 2021. dirección: https://ecuador.unir.net/actualidad-unir/orientacion-vocacional/?utm_source=chatgpt.com
- [71] S. S. Iyengar y M. R. Lepper, “When choice is demotivating: Can one desire too much of a good thing?” *Journal of Personality and Social Psychology*, vol. 79, n.º 6, págs. 995-1006, 2000. DOI: 10.1037/0022-3514.79.6.995
- [72] D. Bawden y L. Robinson, “The dark side of information: Overload, anxiety and other paradoxes and pathologies,” *Journal of Information Science*, vol. 35, n.º 2, págs. 180-191, 2009. DOI: 10.1177/0165551508095781
- [73] F. Ricci, L. Rokach y B. Shapira, “Recommender Systems: Introduction and Challenges,” en *Recommender Systems Handbook*, Springer, 2015, págs. 1-34. DOI: 10.1007/978-1-4899-7637-6_1
- [74] A. J. Head y M. B. Eisenberg, “How students engage with news: Five takeaways for educators, journalists, and librarians,” *First Monday*, vol. 25, n.º 3, 2020. DOI: 10.5210/fm.v25i3.10126
- [75] World Economic Forum, *Future of Jobs Report 2023*, 2023. dirección: <https://www.weforum.org/reports/future-of-jobs-report-2023>
- [76] D. J. Deming, “The growing importance of social skills in the labor market,” *Quarterly Journal of Economics*, vol. 132, n.º 4, págs. 1593-1640, 2017. DOI: 10.1093/qje/qjx022
- [77] C. Succi y M. Canovi, “Soft skills to enhance graduate employability: Comparing students and employers’ perceptions,” *Studies in Higher Education*, vol. 45, n.º 9, págs. 1834-1847, 2020. DOI: 10.1080/03075079.2019.1585420
- [78] OECD, *OECD Skills Outlook 2019: Thriving in a Digital World*, Paris, 2019. DOI: 10.1787/df80bc12-en
- [79] S. McGuinness, K. Pouliakas y P. Redmond, “Skills mismatch: Concepts, measurement and policy approaches,” *Journal of Economic Surveys*, vol. 32, n.º 4, págs. 985-1015, 2018. DOI: 10.1111/joes.12254
- [80] CEDEFOP, “Matching skills and jobs in Europe: Insights from Cedefop’s European skills and jobs survey,” European Centre for the Development of Vocational Training, inf. téc., 2020. dirección: https://www.cedefop.europa.eu/files/3075_en.pdf
- [81] International Labour Organization, *Global Employment Trends for Youth 2020: Technology and the future of jobs*, 2020. dirección: https://www.ilo.org/global/publications/books/WCMS_737648/lang--en/index.htm

- [82] A. G. Watts y R. G. Sultana, "Career guidance policies in 37 countries: Contrasts and common themes," en *International handbook of career guidance*, Springer, 2004, págs. 351-366. DOI: 10.1007/978-1-4020-6230-8_17
- [83] S. Subhash y E. A. Cudney, "Gamified learning in higher education: A systematic review of the literature," *Computers in Human Behavior*, vol. 87, págs. 192-206, 2018. DOI: 10.1016/j.chb.2018.05.028
- [84] T. Hooley, "Career guidance in a digital world: Technologies, ethics and risks," *British Journal of Guidance & Counselling*, vol. 49, n.º 2, págs. 151-164, 2021. DOI: 10.1080/03069885.2020.1828525
- [85] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel y R. Koper, "Recommender systems in technology enhanced learning," *International Journal of Learning Technology*, vol. 5, n.º 4, págs. 409-423, 2011. DOI: 10.1504/IJLT.2010.038082
- [86] H. Drachsler, K. Verbert, O. C. Santos y N. Manouselis, "Panorama of recommender systems to support learning," *Springer Handbook of Learning Analytics*, págs. 407-425, 2015. DOI: 10.1007/978-1-4939-7131-2_21
- [87] M. López-Nores, J. García-Duque, Y. Blanco-Fernández y J. García-Duque, "Educational recommender systems and career guidance," *Computers in Human Behavior*, vol. 104, págs. 106-122, 2020. DOI: 10.1016/j.chb.2019.106122
- [88] C. Fadel, W. Holmes y M. Bialik, "Artificial intelligence in education: Promises and implications for teaching and learning. The Center for Curriculum Redesign, Boston, MA," *Journal of Computer Assisted Learning*, vol. 14, pág. 228, 4 2019, ISSN: 02664909. dirección: https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promise_and_Implications_for_Teaching_and_Learning
- [89] A. Følstad y P. B. Brandtzæg, "Chatbots and the new world of HCI," *Interactions*, vol. 24, n.º 4, págs. 38-42, 2017. DOI: 10.1145/3085558
- [90] E. Kasneci y et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, pág. 102274, 2023. DOI: 10.1016/j.lindif.2023.102274
- [91] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems*, vol. 2020-December, mayo de 2020, ISSN: 10495258. dirección: <https://arxiv.org/pdf/2005.11401>
- [92] R. Caballar y C. Stryker, *¿Qué es un motor de recomendación? | IBM*, jun. de 2024. dirección: <https://www.ibm.com/es-es/think/topics/recommendation-engine>
- [93] Databricks, *What is a Dataset? | Databricks*, 2023. dirección: <https://www.databricks.com/glossary/what-is-dataset>
- [94] A. Badman y M. Kosinski, *What is a Dataset? | IBM*, dic. de 2024. dirección: <https://www.ibm.com/think/topics/dataset>
- [95] I. Belcic y C. Stryker, *What Is Supervised Learning? | IBM*, dic. de 2024. dirección: <https://www.ibm.com/think/topics/supervised-learning>
- [96] N. Nazareth e Y. V. R. Reddy, "Financial applications of machine learning: A literature review," *Expert Systems with Applications*, vol. 219, jun. de 2023, ISSN: 09574174. DOI: 10.1016/j.eswa.2023.119640

- [97] Interactivechaos.com, *Conjunto de datos de prueba* / *Interactive Chaos*. dirección: <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/conjunto-de-datos-de-prueba>
- [98] Domino, *What is a Feature in Machine Learning and Data Science?* / *Domino*, 2024. dirección: <https://domino.ai/data-science-dictionary/feature>
- [99] J. P. Murel, *What is a feature engineering?* / *IBM*, ene. de 2024. dirección: <https://www.ibm.com/think/topics/feature-engineering>
- [100] I. W. Studio, *Métricas de evaluación - Documentación de IBM*, jul. de 2025. dirección: <https://www.ibm.com/docs/es/ws-and-kc?topic=openscale-evaluation-metrics>
- [101] K. Järvelin y J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems*, vol. 20, n.º 4, págs. 422-446, 2002. DOI: 10.1145/582415.582418
- [102] T.-Y. Liu, "Learning to Rank for Information Retrieval," *Foundations and Trends in Information Retrieval*, vol. 3, n.º 3, págs. 225-331, 2009. DOI: 10.1561/15000000016
- [103] C. J. C. Burges, "From RankNet to LambdaRank to LambdaMART: An Overview," Microsoft Research, inf. téc. MSR-TR-2010-82, 2010. dirección: <https://www.microsoft.com/en-us/research/publication/from-ranknet-to-lambdarank-to-lambdamart-an-overview/>
- [104] S. Raschka, J. Patterson y C. Nolet, "Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence," *Information 2020, Vol. 11, Page 193*, vol. 11, pág. 193, 4 abr. de 2020, ISSN: 2078-2489. DOI: 10.3390/INF011040193 dirección: <https://www.mdpi.com/2078-2489/11/4/193/htm%20https://www.mdpi.com/2078-2489/11/4/193>
- [105] "Array programming with NumPy," *Nature 2020 585:7825*, vol. 585, págs. 357-362, 7825 sep. de 2020, ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2 dirección: <https://www.nature.com/articles/s41586-020-2649-2>
- [106] A. Mozzillo, L. Zecchini, L. Gagliardelli, A. Aslam, S. Bergamaschi y G. Simonini, "Evaluation of Dataframe Libraries for Data Preparation on a Single Machine," *Advances in Database Technology - EDBT*, vol. 28, págs. 337-349, 2 nov. de 2024. DOI: 10.48786/edbt.2025.27 dirección: <http://arxiv.org/abs/2312.11122%20http://dx.doi.org/10.48786/edbt.2025.27>
- [107] N. Abdennur et al., "Bioframe: operations on genomic intervals in Pandas dataframes," *Bioinformatics*, vol. 40, btac088, 2 feb. de 2024, ISSN: 13674811. DOI: 10.1093/BIOINFORMATICS/BTAE088 dirección: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10903647/>
- [108] A. Goscinski et al., "scikit-matter : A Suite of Generalisable Machine Learning Methods Born out of Chemistry and Materials Science," *Open Research Europe*, vol. 3, pág. 81, 2023, ISSN: 27325121. DOI: 10.12688/OPENRESEUROPE.15789.2 dirección: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10792272/>

- [109] G. Stoet, “PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments,” *Teaching of Psychology*, vol. 44, págs. 24-31, 1 ene. de 2017, ISSN: 15328023. DOI: 10.1177/0098628316677643; WEBSITE: WEBSITE: SAGE; JOURNAL: JOURNAL: TOPA; WGROUP: STRING: PUBLICATION dirección: /doi/pdf/10.1177/0098628316677643?download=true
- [110] J. Kim, U. Gabriel y P. Gyga, “Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task,” *PLoS ONE*, vol. 14, e0221802, 9 sep. de 2019, ISSN: 19326203. DOI: 10.1371/JOURNAL.PONE.0221802 dirección: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6726137/>
- [111] T. Chen y C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” págs. 785-794, 2016. DOI: 10.1145/2939672.2939785 dirección: <https://doi.org/10.1145/2939672.2939785>
- [112] J. Pan, G. Li et al., “Survey of Vector Database Management Systems,” *VLDB Journal (preprint on arXiv:2310.14021)*, 2024. dirección: <https://dbgroup.cs.tsinghua.edu.cn/lvgl/papers/vldb2024-vectordb.pdf>
- [113] Pinecone Systems, Inc. “Architecture — Pinecone Docs.” Accessed 2025-10-13. dirección: <https://docs.pinecone.io/guides/get-started/database-architecture>
- [114] Pinecone Systems, Inc. “What is a Vector Database & How Does it Work?” Dirección: <https://www.pinecone.io/learn/vector-database/>
- [115] MongoDB, Inc. “What Is NoSQL? NoSQL Databases Explained.” dirección: <https://www.mongodb.com/resources/basics/databases/nosql-explained>
- [116] MongoDB, Inc. “Documents — MongoDB Manual.” dirección: <https://www.mongodb.com/docs/manual/core/document/>
- [117] MongoDB, Inc. “Document Database — NoSQL Basics.” dirección: <https://www.mongodb.com/resources/basics/databases/document-databases>
- [118] Amazon Web Services. “What is Amazon DocumentDB (with MongoDB compatibility)?” Dirección: <https://docs.aws.amazon.com/documentdb/latest/developerguide/what-is.html>
- [119] I. Postman, *What is Postman?* <https://www.postman.com/product/what-is-postman/>, Último acceso: 2025, 2025.
- [120] I. Postman, *Test your API using the Collection Runner*, <https://learning.postman.com/docs/collections/running-collections/intro-to-collection-runs/>, Último acceso: 2025, 2025.
- [121] S. U. Noble, *Algorithms of oppression: How search engines reinforce racism*. NYU Press, 2018.
- [122] L. Floridi y M. Taddeo, “What is data ethics?” *Philosophical Transactions of the Royal Society A*, vol. 374, n.º 2083, pág. 20160360, 2016. DOI: 10.1098/rsta.2016.0360
- [123] F. Doshi-Velez y B. Kim, “Towards a rigorous science of interpretable machine learning,” *arXiv preprint*, 2017. eprint: 1702.08608.

- [124] A. -. A. S. R. of Explainable Artificial Intelligence Models y Methods, “A Systematic Review of Explainable Artificial Intelligence Models and Methods,” *Elsevier Digital Library*, 2023, “Explainable Artificial Intelligence (XAI) is a term that refers to Artificial Intelligence (AI) that can provide explanations for their decision or predictions.”. dirección: <https://www.sciencedirect.com/science/article/pii/S277266222300070X>
- [125] D. Gunning et al., “Four Principles of Explainable Artificial Intelligence,” *NIST Internal Report 8312*, 2021, Outlines four principles for explainable AI systems: explanation, meaningfulness, accuracy of explanation, knowledge limits. dirección: <https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf>
- [126] M. Ahmad et al., “Explainable Artificial Intelligence (XAI): Current Status and Future Potential,” *Frontiers in Artificial Intelligence*, 2022, “Explainable artificial intelligence (XAI) is a young, rapidly evolving, and exciting field.”. dirección: <https://www.ncbi.nlm.nih.gov/articles/PMC10853303/>
- [127] C. M. U. S. Blog. “What is Explainable AI?” “Explainable artificial intelligence is a powerful tool in answering critical How? and Why? questions about AI systems ...”. dirección: <https://insights.sei.cmu.edu/blog/what-is-explainable-ai/>
- [128] K. Babitz, *Bases de datos SQL vs NoSQL: Diferencias clave e ideas prácticas | DataCamp*, ene. de 2025. dirección: <https://www.datacamp.com/es/blog/sql-vs-nosql-databases>
- [129] Amazon.com, *Tipos de bases de datos NoSQL: Cómo elegir una base de datos NoSQL de AWS*. dirección: <https://docs.aws.amazon.com/whitepapers/latest/choosing-an-aws-nosql-database/types-of-nosql-databases.html>
- [130] Arsys, *Bases de datos NoSQL: qué son, tipos y ventajas | Blog de Arsys*, dic. de 2023. dirección: <https://www.arsys.es/blog/bases-de-datos-nosql-que-son-tipos-y-ventajas>
- [131] StackScale, *Bases de datos NoSQL: características y tipos*, ene. de 2023. dirección: <https://www.stackscale.com/es/blog/bases-de-datos-nosql/>
- [132] Chroma Inc. “Chroma – the open-source search and retrieval database for AI applications.” Acceso el 24 oct 2025. dirección: <https://www.trychroma.com/>
- [133] Chroma Inc. “Chroma Cloud: simple usage-based pricing.” Acceso el 24 oct 2025. dirección: <https://www.trychroma.com/pricing>
- [134] Pinecone Systems Inc. “Pinecone Pricing – Start free, scale effortlessly.” Acceso el 24 oct 2025. dirección: <https://www.pinecone.io/pricing/>
- [135] Weaviate B.V. “The AI-Native, Open Source Vector Database – Weaviate Platform.” Acceso el 24 oct 2025. dirección: <https://weaviate.io/platform>
- [136] Aloa AI, “Pinecone vs Weaviate vs Chroma – The definitive 2025 vector database comparison,” 2025, Acceso el 24 oct 2025. dirección: <https://aloe.co/ai/comparisons/vector-database-comparison/pinecone-vs-weaviate-vs-chroma>
- [137] Meta AI Research, *Faiss: A library for efficient similarity search and clustering of dense vectors*, Acceso el 24 oct 2025, 2025. dirección: <https://faiss.ai/index.html>
- [138] Qdrant Team. “Qdrant Vector Database – Efficient, scalable, fast.” Acceso el 24 oct 2025. dirección: <https://qdrant.tech/qdrant-vector-database/>

- [139] I. Belcic y C. Stryker. “RAG vs. Fine-Tuning.” Acceso el 24 oct 2025, IBM. dirección: <https://www.ibm.com/think/topics/rag-vs-fine-tuning>
- [140] L. MacDonald. “RAG vs Fine Tuning: How to Choose the Right Method.” Acceso el 24 oct 2025, Monte Carlo Data. dirección: <https://www.montecarlodata.com/blog-rag-vs-fine-tuning/>
- [141] F. Wunderlich. “Fine-Tuning vs. RAG: Understanding the Difference.” Acceso el 24 oct 2025, FinetuneDB. dirección: <https://finetunedb.com/blog/fine-tuning-vs-rag/>
- [142] J. Erickson. “RAG vs. Fine-Tuning: How to Choose.” Acceso el 24 oct 2025, Oracle. dirección: <https://www.oracle.com/artificial-intelligence/generative-ai/retrieval-augmented-generation-rag/rag-fine-tuning/>
- [143] Cohere, *Embed | Secure AI Retrieval | Cohere*, 2025. dirección: <https://cohere.com/embed>
- [144] Cohere, *Cohere SDK Cloud Platform Compatibility | Cohere*, 2021. dirección: <https://docs.cohere.com/v1/docs/cohere-works-everywhere>
- [145] OpenAI, *Compare models - OpenAI API*, 2025. dirección: <https://platform.openai.com/docs/models/compare?model=text-embedding-3-small>
- [146] G. A. for Developers, *Embeddings | Gemini API | Google AI for Developers*, 2025. dirección: <https://ai.google.dev/gemini-api/docs/embeddings?hl=es-419>
- [147] Stevia Putri. “A Complete Guide to Claude Pricing in 2025.” Consultado el 24 oct 2025. dirección: <https://www.eesel.ai/blog/claude-pricing>
- [148] Anthropic PBC. “Claude 3 Haiku: Our Fastest Model Yet.” Consultado el 24 oct 2025. dirección: <https://www.anthropic.com/news/claude-3-haiku>
- [149] Mistral AI Team. “AI in Abundance: Introducing a Free API, Improved Pricing, and a New Mistral Small.” Consultado el 24 oct 2025. dirección: <https://mistral.ai/news/september-24-release>
- [150] Cohere Inc. “Pricing – Secure and Scalable Enterprise AI.” Consultado el 24 oct 2025. dirección: <https://cohere.com/pricing>
- [151] Google AI. “Gemini Developer API Pricing.” Consultado el 24 oct 2025. dirección: <https://ai.google.dev/gemini-api/docs/pricing>
- [152] J. Almeida. “What are the Best Languages for Machine Learning in 2025?” Consultado en octubre de 2025. dirección: <https://distantjob.com/blog/best-language-machine-learning/>
- [153] A. Marketing. “Python vs R vs Julia: Best Language for Machine Learning 2025.” Consultado en octubre de 2025. dirección: <https://ai.devtheworld.jp/posts/python-vs-r-vs-julia-ml-2025/>
- [154] J. Delua. “Supervised versus unsupervised learning: What’s the difference?” Consultado en octubre de 2025. dirección: <https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning>
- [155] R. Qayyum. “Computational Complexity of Machine Learning Algorithms.” Consultado en octubre de 2025. dirección: <https://rafayqayyum.medium.com/computational-complexity-of-machine-learning-algorithms-254c275de84>

- [156] E. Dorfman. “How Much Data Is Required for Machine Learning?” Consultado en octubre de 2025. dirección: <https://postindustria.com/how-much-data-is-required-for-machine-learning/>
- [157] GeeksforGeeks Editorial Team. “Gradient Boosting vs Random Forest.” Consultado en octubre de 2025. dirección: <https://www.geeksforgeeks.org/machine-learning/gradient-boosting-vs-random-forest/>
- [158] P. Aznar. “What is the difference between Extra Trees and Random Forest?” Consultado en octubre de 2025. dirección: <https://quantdare.com/what-is-the-difference-between-extra-trees-and-random-forest/>
- [159] T. DigitalDefynd. “10 Pros & Cons of Decision Tree Algorithms [2025].” Consultado en octubre de 2025. dirección: <https://digitaldefynd.com/IQ/decision-tree-algorithms-pros-cons/>
- [160] H. Idrees. “Gradient Boosting vs. Random Forest: Which Ensemble Method Should You Use?” Consultado en octubre de 2025. dirección: <https://medium.com/@hassaanidrees7/gradient-boosting-vs-random-forest-which-ensemble-method-should-you-use-9f2ee294d9c6>
- [161] C. Dilmegani. “GraphQL vs. REST: Top 4 Advantages & Disadvantages.” Consultado en octubre de 2025. dirección: <https://research.aimultiple.com/graphql-vs-rest/>
- [162] Y. Lin. “GraphQL vs REST API: Which is Better for Your Project in 2025?” Consultado en octubre de 2025. dirección: <https://api7.ai/blog/graphql-vs-rest-api-comparison-2025>
- [163] K. Inc. “GraphQL vs REST: Key Similarities and Differences Explained.” Consultado en octubre de 2025. dirección: <https://konghq.com/blog/learning-center/graphql-vs-rest>
- [164] R. Learning. “What Languages Should API Developers Learn? Best Programming Languages for an API Development Career.” Consultado en octubre de 2025. dirección: <https://www.refontelearning.com/blog/what-languages-should-api-developers-learn>
- [165] F. Doglio. “The 5 Best Backend Development Languages to Master (2025).” Consultado en octubre de 2025. dirección: <https://roadmap.sh/backend/languages>
- [166] U. Technology. “How to: Best Languages to Develop Efficient REST APIs.” Consultado en octubre de 2025. dirección: <https://www.unimedia.tech/best-languages-to-develop-efficient-rest-apis/>
- [167] S. Islam. “Express vs Koa vs Fastify vs NestJS vs Hono: Choosing the Right Node.js Framework.” Consultado en octubre de 2025. dirección: <https://medium.com/@khanshahid9283/express-vs-koa-vs-fastify-vs-nestjs-vs-hono-choosing-the-right-node-js-framework-17a56a533d29>
- [168] E. Team. “Node.js Frameworks Roundup 2024—Elysia, Hono, Nest, Encore — Which should you pick?” Consultado en octubre de 2025. dirección: <https://dev.to/encore/nodejs-frameworks-roundup-2024-elysia-hono-nest-encore-which-should-you-pick-19oj>

- [169] F. Hicks. “Best Node.js Web Framework – Which One Should You Choose?” Consultado en octubre de 2025. dirección: <https://www.aegissofttech.com/insights/best-node-js-web-frameworks/>
- [170] Universidad del Valle de Guatemala. “Carreras de la Universidad del Valle de Guatemala.” dirección: <https://www.uvg.edu.gt/academico/carreras/>
- [171] TuSalario.com. “TuSalario.com.” dirección: <https://tusalarario.org/guatemala/Tucarrera/guatemala-ocupacion-y-salario?occupation=ingenier%C3%ADa+civil>
- [172] Salary.com. “Salary.com.” dirección: <https://www.salary.com/research/salary>
- [173] M. Palczyńska y K. Świst, “Personality, cognitive skills and life outcomes: evidence from the Polish follow-up study to PIAAC,” *Large-Scale Assessments in Education*, vol. 6, págs. 1-23, 1 dic. de 2018, ISSN: 21960739. DOI: 10.1186/S40536-018-0056-Z/TABLES/12 dirección: <https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/s40536-018-0056-z>
- [174] B. Rammstedt y D. Danner, “DIRECTORATE FOR EDUCATION AND SKILLS Beyond Literacy: The Incremental Value of Non-Cognitive Skills OECD Education Working Paper No. 311,” mar. de 2024. dirección: <http://www.oecd.org/termsandconditions>.

11.1. Resultados

11.1.1. Resultados módulo de información

Detalles sobre la información recopilada

Colección de carreras (Careers) Colección de tarjetas (Cards) Esta colección es el núcleo de la base de conocimiento, ya que contiene información detallada y estructurada sobre cada una de las 33 carreras profesionales ofrecidas en la Universidad del Valle de Guatemala. Cada documento de esta colección representa una única carrera e incluye datos académicos, profesionales y de competencias. **Descripción de los campos**

Cantidad recolectada: 53+ tarjetas (20 testimonios, 33 carreras y las generadas por los

```
"_id": ObjectId(),
```

Colección de reglas de puntuación (Scoring rules)

```
"facultad": string,
```

Descripción de los campos

```
"duracion": number,
```

Colección de etiquetas (Tags)

```
"areas_de_desarrollo_potencial": Array<{ "area": string, "descripcion": string }>,
```

Descripción de los campos

```
"año": { "semestre": [ { id: ObjectId(), name: string } ] }
```

Cantidad recolectada: 26 etiquetas

```
"perfil_del_egresado": string,
```

Colección de ítems de test (Test Items)

```
"salario_maximo": number,
```

Descripción de los campos

```
"tags": Array<{ "tag": ObjectId(), "name": string, "score": number }>
```

Cantidad recolectada: 47 preguntas

Consultas realizadas en la prueba del objetivo de base de datos Descripción de los campos

- **_id**: identificador único para el documento de la carrera.
- **nombre_carrera**: el nombre oficial de la carrera.
- **facultad**: la facultad a la que pertenece la carrera.
- **descripcion**: un resumen detallado de lo que trata la carrera.
- **duración**: la duración de la carrera en años.
- **empleabilidad**: una estimación del nivel de empleabilidad para los graduados.
- **areas_de_desarrollo_potencial**: posibles campos en los que un profesional de esta carrera puede especializarse.
- **plan_de_estudio**: una estructura anidada que contiene los IDs de los cursos para cada semestre y año.
- **perfil_del_egresado**: describe las capacidades y oportunidades del profesional graduado.
- **salario_minimo** y **salario_maximo**: rango salarial estimado para un profesional en esta área.
- **tags**: un arreglo de etiquetas que asocian la carrera con habilidades, intereses y rasgos de personalidad, cada una con un puntaje (**score**) que indica su relevancia. Estos tags son clave para el sistema de recomendación.

Esta estructura detallada permite al sistema tener un conocimiento profundo de cada carrera, lo que facilita consultas precisas y la generación de recomendaciones relevantes basadas en la alineación entre los atributos de la carrera y el perfil del usuario.

Cantidad recolectada: 33 carreras

Colección de cursos (Courses)

Para complementar la información de las carreras, se ha creado una colección de cursos que cataloga todas las materias que forman parte de los planes de estudio. Esta colección es esencial para entender la estructura académica de cada carrera y las dependencias entre cursos.

```
{
  "_id": ObjectId(),
  "nombre": string,
  "prerequisitos": [ObjectId()]
}
```

Descripción de los campos

- **_id**: identificador único del curso en la base de datos.
- **nombre**: el nombre completo de la materia.
- **prerequisitos**: un arreglo que contiene los identificadores (**_id**) de los cursos que son necesarios para poder tomar esta materia. Si el arreglo está vacío, el curso no tiene prerequisites.

Esta colección permite construir el mapa curricular completo de cada carrera y visualizar las rutas de aprendizaje que un estudiante debe seguir.

Cantidad recolectada: 897 cursos

Colección de testimonios de egresado (Testimonies) Una de las colecciones de la base de conocimiento es la de **testimonios**. Esta colección recopila las experiencias de egresados de distintas carreras de la Universidad del Valle de Guatemala. El objetivo de estos testimonios es ofrecer una visión auténtica y personal sobre lo que implica estudiar una carrera específica, los desafíos que se pueden encontrar y las oportunidades profesionales que surgen tras la graduación.

Cantidad recolectada: 20 testimonios

```
{
  "_id": ObjectId(),
  "name": string,
  "description": string,
  "created_at": date,
  "updated_at": date
}
```

Colección de tarjetas (Cards)

```
{
  "_id": ObjectId(),
  "type": string,
  "display_data": Object{},
  "reference": {
    "collection": string,
    "id": ObjectId()
  },
  "created_at": date,
  "updated_at": date
}
```

Descripción de los campos

Cada campo del documento JSON tiene un propósito específico para la recolección y consulta de la información:

- **`__id`**: identificador único del documento en la base de datos, asignado por MongoDB.
- **`type`**: define el tipo de información que contiene el documento. En este caso, es un **`testimony`**. Otros tipos pueden incluir descripciones de carrera, preguntas interesantes, etc.
- **`display_data`**: objeto que contiene la información principal que se mostrará en la tarjeta.
- **`reference`**: objeto que vincula este documento con otras colecciones en la base de datos, manteniendo la integridad relacional.
- **`created_at`** y **`updated_at`**: marcas de tiempo que registran cuándo fue creado y actualizado el documento, respectivamente.

Esta estructura nos permite guardar información de diferentes tipos, sin la necesidad de limitarnos a tener una estructura rígida, sino flexible y adaptable. Está diseñada para permitirnos realizar consultas complejas que relacionen diferentes piezas de información, lo cual es fundamental para el modelo de recomendación.

Cantidad recolectada: 53+ tarjetas (20 testimonios, 33 carreras y las generadas por los usuarios)

Colección de reglas de puntuación (scoring rules)

Esta colección es fundamental para la evaluación psicométrica del perfil de usuario. Contiene las reglas y la estructura necesarias para determinar el puntaje de las respuestas de los usuarios a los cuestionarios de intereses y personalidad, como el test RIASEC.

```
{
  "_id": ObjectId(),
  "section": string,
  "description": string,
  "traits": Array<
    {
      "name": string,
      "code": string,
      "description": string,
      "questions": Array<string>,
      "reverse_scored": Array<string>
    }
  >
}
```

Descripción de los campos

- **_id**: identificador único para la regla de puntuación.
- **section**: el nombre de la sección o test al que pertenecen las reglas (p. ej., **riasec**).
- **description**: una breve descripción del test o la sección.
- **traits**: un arreglo de objetos, donde cada objeto representa un rasgo a medir.
 - **name**: el nombre del rasgo (p. ej., **realistic**).
 - **code**: el código o abreviatura del rasgo (p. ej., **R**).
 - **description**: una descripción del rasgo.
 - **questions**: un arreglo de identificadores de las preguntas que miden este rasgo.
 - **reverse_scored**: un arreglo que contiene las preguntas cuya puntuación debe ser invertida.

Esta estructura permite al sistema procesar las respuestas de los usuarios de manera automatizada, calcular las puntuaciones para cada rasgo y, finalmente, construir un perfil psicométrico preciso que se utilizará para las recomendaciones de carrera.

Colección de Etiquetas (Tags) Colección de cursos (Courses) La colección de etiquetas es un componente central que conecta todas las demás colecciones de la base de conocimiento. Cada etiqueta representa un concepto, habilidad, interés o rasgo de personalidad que puede ser utilizado para describir tanto a las carreras y testimonios como a los perfiles de usuario. **Descripción de los campos**

```
Colección de testimonios de egresado (Testimonies)
  "_id": ObjectId(),
Colección de tarjetas (Cards)
  "description": string
Descripción de los campos
```

Cantidad recolectada: 53+ tarjetas (20 testimonios, 33 carreras y las generadas por los usuarios) **Descripción de los campos** **Colección de reglas de puntuación (Scoring Rules)**

- **Descripción de los campos**
- **name**: el nombre de la etiqueta, que funciona como un identificador legible (p. ej., **openness, biology, realistic**).
- **Colección de etiquetas (Tags)**

Descripción de los campos Estas etiquetas son la base para el sistema de recomendación, ya que permiten crear un lenguaje común para comparar las características de las carreras con los intereses y aptitudes de los usuarios. **Cantidad recolectada:** 26 etiquetas **Cantidad recolectada:** 26 etiquetas **Colección de ítems de test (Test Items)**

Descripción de los campos

Cantidad recolectada: 47 preguntas

Consultas realizadas en la prueba del objetivo de base de datos Test de personalidad (Big Five):

```
"id": string,  
  Escala de medición: Likert de 7 puntos.  
"type": string,  
  Opciones de escala:  
"options": Array<{ "id": string, "label": string }>,  
  Ítems por rasgo:  
"meta": Object{  
  Apertura a la experiencia (Openness):
```

Conciencia (Conscientiousness): Descripción de los campos Extraversión (Extraversion):

- **Amabilidad (Agreeableness):**
- **id:** un identificador legible y único para la pregunta (p. ej., `metadata_gender`).
- **section:** la sección del test a la que pertenece el ítem (p. ej., `metadata_riasec`).
- **type:** el tipo de pregunta, que define cómo se mostrará en la interfaz (p. ej., `radio`, `likert`).
- **prompt:** el texto de la pregunta que se le presenta al usuario.
- **options:** un arreglo de objetos que define las posibles respuestas para la pregunta.
- **correctAnswer:** utilizado para preguntas con una respuesta correcta; para tests psicométricos, suele ser nulo.
- **meta:** un objeto para almacenar metadatos adicionales sobre la pregunta.

La estandarización de las preguntas en esta colección permite una construcción flexible y modular de los cuestionarios, así como un procesamiento consistente de las respuestas.

Cantidad recolectada: 47 preguntas

Consultas realizadas en la Prueba del Objetivo de Base de Datos

```
[OK] Facultad a la que pertenece la carrera Ingeniería Biomédica  
operación: equals  
esperado: Facultad de Ingenieria  
encontrado: Facultad de Ingenieria  
[OK] Duración de la carrera de Ingeniería en Biotecnología Industrial  
operación: equals  
esperado: 5
```

Figura 29. Resultados de la ejecución de la prueba del objetivo de la base de datos

```

python3 dataTesting.py

=====
VALIDACIÓN DE BASE DE CONOCIMIENTO
Fecha: 2025-10-30T15:22:25
DB: mirai | Collection: careers
=====

[1] Integridad estructural
  - Documentos totales      : 33
  - Documentos válidos     : 33
  - Precisión (integridad)  : 100.0%
  - No se encontraron documentos incompletos.

[2] Precisión de consulta (Q/A controladas)
  - Pruebas totales        : 30
  - Pruebas correctas      : 30
  - Precisión (consultas)  : 100.0%
  - Detalle de pruebas:

```

Nota. Elaboración propia.

```

    encontrado: 5
[OK] Facultad a la que pertenece la carrera Ingeniería en Ciencias de
↳ Alimentos
    operación: equals
    esperado: Facultad de Ingenieria
    encontrado: Facultad de Ingenieria
[OK] Duración de la carrera de Ingeniería Civil Arquitectónica
    operación: equals
    esperado: 6
    encontrado: 6
[OK] Contiene curso en segundo año, primer semestre de la carrera Ingeniería
↳ en Ciencia de la Administración
    operación: contains
    esperado: Contabilidad de Costos
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9d43'), 'name':
↳ 'Contabilidad de Costos'}
[OK] Contiene curso en tercer año, primer semestre de la carrera Ingeniería
↳ en Ciencia de la Computación y Tecnologías de la Información
    operación: contains
    esperado: Bases de Datos 1
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9ceb'), 'name': 'Bases
↳ de Datos 1'}
[OK] Contiene curso en año 3, primer semestre de Licenciatura en Sistemas de
↳ Información Computacional
    operación: contains
    esperado: Bases de Datos 2
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9cec'), 'name': 'Bases
↳ de Datos 2'}
[OK] Facultad a la que pertenece la carrera Ingeniería Electrónica
    operación: equals
    esperado: Facultad de Ingenieria
    encontrado: Facultad de Ingenieria
[OK] Duración de la carrera de Ingeniería Industrial
    operación: equals
    esperado: 5

```

```

    encontrado: 5
[OK] Contiene curso en año 5, segundo semestre de la carrera Ingeniería Civil
    operación: contains
    esperado: Diseño de Obras Hidraulicas y Obras Viales
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9d82'), 'name': 'Diseño
    ↪ de Obras Hidraulicas y Obras Viales'}
[OK] Contiene curso en año 1, segundo semestre de Licenciatura en Biología
    operación: contains
    esperado: Biologia General
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9cf8'), 'name':
    ↪ 'Biologia General'}
[OK] Facultad a la que pertenece la carrera Ingeniería Mecatrónica
    operación: equals
    esperado: Facultad de Ingenieria
    encontrado: Facultad de Ingenieria
[OK] Contiene curso en año 3, primer semestre de Ingeniería Mecánica
    operación: contains
    esperado: Termodinamica 2
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9fd2'), 'name':
    ↪ 'Termodinamica 2'}
[OK] Facultad a la que pertenece la carrera Licenciatura en Comunicación
    ↪ Estratégica
    operación: equals
    esperado: Bridge Business School
    encontrado: Bridge Business School
[OK] Duración de la carrera Ingeniería Mecánica Industrial
    operación: equals
    esperado: 6
    encontrado: 6
[OK] Facultad a la que pertenece la carrera Licenciatura en Administración de
    ↪ Empresas
    operación: equals
    esperado: Bridge Business School
    encontrado: Bridge Business School
[OK] Facultad a la que pertenece la carrera Licenciatura en Arquitectura
    operación: equals
    esperado: Escuela de Arquitectura
    encontrado: Escuela de Arquitectura
[OK] Contiene curso en año 3, segundo semestre de la carrera Ingeniería
    ↪ Química
    operación: contains
    esperado: Metodos Numericos 1
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9ed4'), 'name': 'Metodos
    ↪ Numericos 1'}
[OK] Contiene curso en año 4, primer semestre de Ingeniería Química
    ↪ Industrial
    operación: contains
    esperado: Manufactura en Ingenieria Quimica
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9eb2'), 'name':
    ↪ 'Manufactura en Ingenieria Quimica'}
[OK] Contiene curso en año 3, segundo semestre de International Marketing and
    ↪ Business Analytics
    operación: contains
    esperado: Data Mining & Machine Learning

```

```

    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9d58'), 'name': 'Data
    ↪ Mining & Machine Learning'}
[OK] Contiene curso en año 4, primer semestre de Licenciatura en Física
    operación: contains
    esperado: Mecanica Cuantica 1
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9ec4'), 'name':
    ↪ 'Mecanica Cuantica 1'}
[OK] Duración de la carrera Licenciatura en Psicología
    operación: equals
    esperado: 5
    encontrado: 5
[OK] Facultad a la que pertenece la carrera Licenciatura en Biotecnología
    ↪ Molecular
    operación: equals
    esperado: Facultad de Ciencias y Humanidades
    encontrado: Facultad de Ciencias y Humanidades
[OK] Nombre completo de la carrera Licenciatura en Bioquímica y Microbiología
    operación: equals
    esperado: Licenciatura en Bioquimica y Microbiologia
    encontrado: Licenciatura en Bioquimica y Microbiologia
[OK] Duración de la carrera Licenciatura en Química Farmacéutica
    operación: equals
    esperado: 5
    encontrado: 5
[OK] Contiene curso en año 4, segundo semestre de Licenciatura en Arqueología
    operación: contains
    esperado: Epigrafia Maya
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9dbc'), 'name':
    ↪ 'Epigrafia Maya'}
[OK] Facultad a la que pertenece la carrera Licenciatura en Antropología
    operación: equals
    esperado: Facultad de Ciencias Sociales
    encontrado: Facultad de Ciencias Sociales
[OK] Contiene curso en año 4, primer semestre de Licenciatura en Química
    operación: contains
    esperado: Quimica cuantica
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9f4f'), 'name': 'Quimica
    ↪ cuantica'}
[OK] Contiene curso en año 2, primer semestre de Licenciatura en Matemática
    ↪ Aplicada
    operación: contains
    esperado: Algoritmos y Estructuras de Datos
    encontrado: {'id': ObjectId('68df5ddf3309b8126b3f9cbc'), 'name':
    ↪ 'Algoritmos y Estructuras de Datos'}
[OK] Duración de la carrera Licenciatura en Nutrición
    operación: equals
    esperado: 5
    encontrado: 5

```

Ítems seleccionados por instrumento

Test de personalidad (Big Five):

- **Escala de medición:** Likert de 7 puntos.
 - **Opciones de escala:**
 - Totalmente en Desacuerdo (1)
 - En Desacuerdo (2)
 - Algo en Desacuerdo (3)
 - Neutral (4)
 - Algo de Acuerdo (5)
 - De Acuerdo (6)
 - Totalmente De Acuerdo (7)
- **Ítems por rasgo:**
 - **Apertura a la experiencia (Openness):**
 - Se preocupa mucho
 - Se pone nervioso fácilmente
 - Permanece tranquilo en situaciones tensas **inverso**
 - **Conciencia (Conscientiousness):**
 - Es hablador
 - Es extrovertido, sociable
 - Es reservado **inverso**
 - **Extraversión (Extroversion):**
 - Es original, tiene nuevas ideas
 - Valora las experiencias artísticas y estéticas
 - Tiene una imaginación activa
 - **Amabilidad (Agreeableness):**
 - A veces es grosero con los demás **inverso**
 - Tiene una naturaleza indulgente
 - Es considerado y amable con casi todos
 - **Neuroticismo (Neuroticism):**
 - Hace un trabajo minucioso
 - Tiende a ser perezoso **inverso**
 - Hace las cosas de manera eficiente

Test de intereses vocacionales (RIASEC):

- **Escala de medición:** Likert de 5 puntos.
 - **Opciones de escala:**
 - Totalmente en Desacuerdo (1)
 - En Desacuerdo (2)
 - Neutral (3)
 - De Acuerdo (4)

- Totalmente De Acuerdo (5)

■ **Ítems por tipo vocacional:**

- **Realista (Realistic):**
 - Me gusta construir cosas
 - Me gusta trabajar al aire libre
- **Investigador (Investigative):**
 - Disfruto intentando averiguar cómo funcionan las cosas
 - Me gusta analizar cosas (problemas/situaciones)
- **Artístico (Artistic):**
 - Soy una persona creativa
 - Se me da bien trabajar de forma independiente
- **Social (Social):**
 - Me gusta trabajar en equipo
 - Me gusta intentar ayudar a la gente a resolver sus problemas
- **Emprendedor (Entrepreneurial):**
 - Me gusta liderar
 - Asumo rápidamente nuevas responsabilidades
- **Convencional (Conventional):**
 - Me gusta organizar cosas (archivos, escritorios/oficinas)
 - Presto atención a los detalles

Test de persistencia y consistencia (Grits):

■ **Escala de medición:** Likert de 5 puntos.

- **Opciones de escala:**
 - Muy parecido a mí (5)
 - Bastante parecido a mí (4)
 - Algo parecido a mí (3)
 - Poco parecido a mí (2)
 - Nada parecido a mí (1)

■ **Ítems por grit:**

- **Consistencia (Consistency):**
 - Las nuevas ideas y proyectos a veces me distraen de los anteriores.
 - He estado obsesionado con una idea o proyecto por un corto tiempo, pero luego perdí el interés.
 - A menudo establezco una meta, pero luego decido perseguir una diferente.
 - Tengo dificultad para mantener mi enfoque en proyectos que toman más de unos pocos meses en completarse.

- Perseverancia (Perseverance):
 - Los contratiempos no me desaniman. **inverso**
 - Soy un trabajador arduo. **inverso**
 - Termino todo lo que comienzo. **inverso**
 - Soy diligente. **inverso**

Prueba de habilidades cognitivas (PIAAC):

■ Comprensión lectora (reading comprehension):

1. Ítem 1:

• Inciso:

****¡Bienvenidos a nuestra escuela infantil!****

Esperamos disfrutar de un gran año de diversión, aprendizaje y conocimiento mutuo. Por favor, tómese un momento para revisar las normas de nuestra escuela infantil.

- Por favor, traiga a su hijo/a no más tarde de las 09:00 h.
- Vista a su hijo/a de forma cómoda y traiga una muda de ropa.
- Por favor, no traiga joyas o caramelos. Si su hijo/a cumple años, por favor, hable con su profesor/a sobre alguna chuchería especial para los niños.
- Por favor, traiga a su hijo/a bien vestido, no en pijama.
- El desayuno se servirá hasta las 07:30 h.
- Traiga una manta o almohada pequeñas. Por favor, deje los juguetes en casa.
- La medicación tiene que venir en su envase original, etiquetada, y hay que firmar la hoja de medicación que hay en cada aula.

- **Prompt:** ¿A qué hora, como muy tarde, deben llegar los niños a la escuela infantil?

- **Respuesta correcta:** 09:00 h.

2. Ítem 2:

• Inciso:

****El pan se endurece, pero las galletas saladas se ablandan.****

¿Por qué el pan se endurece y estropea cuando se expone al aire? Parte de la explicación se debe a que pierde humedad. El pan blando corriente tiene entre un ****32 y 38 %**** de humedad. Si dejamos el pan sin envolver y expuesto a los elementos, pierde humedad frente al aire y terminará endureciéndose cuando el nivel de humedad desciende al ****14 %****, aproximadamente.

Al tiempo que la humedad del pan se evapora, se da un proceso

llamado *retrogradación* por el cual la estructura del almidón cambia. Durante la retrogradación, la corteza del pan se ablanda y la parte media del pan se endurece. Además, una parte del almidón se cristaliza. Cuando esto sucede, se produce un endurecimiento del pan al tiempo que se estropea.

Los almidones duros, como el de las galletas saladas, son crujientes porque están elaborados bajo un nivel de humedad extremadamente bajo, normalmente entre el **2 y el 5 %**. Cuando se exponen al aire, las galletas saladas absorben la humedad del aire. Las galletas parecen blandas cuando su nivel de humedad alcanza un **9 %**, aproximadamente.

- **Prompt:** Selecciona para cada afirmación si aplica a Pan, Galleta o Ambos.

- **Opciones:**

- a) Pan
 - b) Galleta
 - c) Ambos

- **Incisos:**

- a) Debe envolverse para mantenerse fresco = Ambos
 - b) Más fresco cuando está blando = Pan
 - c) Le afecta la exposición al aire = Ambos

- **Razonamiento matemático (mathematical reasoning):**

- 1. **Ítem 1:**

- **Inciso:**

Figura 30. PIAAC - Unidad 817: Pregunta 1

El mortero es un material de construcción que se utiliza para aplicar un acabado texturizado que cubra los ladrillos o bloques de cemento de los muros. En el saco se indica la superficie que se puede cubrir al aplicar una capa de grosor medio.



- **Prompt:** ¿Cuántos kg de mezcla de mortero necesita para un muro que mide 5 por 4 metros?
- **Respuesta correcta:** 80 kg.

2. Ítem 2:

- **Inciso:**

Figura 31. PIAAC - Unidad 826: Pregunta 2

Temperaturas en la Cámara frigorífica

Una cámara frigorífica mantiene los alimentos congelados en una planta de comida procesada. La temperatura se mide en grados Celsius (°C) y se muestra en un termómetro digital.

La temperatura en la cámara frigorífica debería mantenerse siempre dentro del rango 20.0 °C a -15.0 °C.



- **Prompt:** La temperatura actual es -15.2°C y desciende 1.5°C. ¿Cuál será la nueva temperatura?
- **Opciones:**
 - a) -13.7°C
 - b) -16.7°C
 - c) -14.7°C

■ Resolución de problemas (problem solving):

- **Inciso:**

Figura 32. PIAAC - Unidad 826: Pregunta 3

Las siguientes comprobaciones son necesarias como parte del proceso de control de calidad del loncheado y empaquetado de embutidos en la planta de comida procesada.

COMPROBACIONES DE CALIDAD: ROAST BEEF LONCHEADO

- Temperatura interna antes del loncheado: 0°C a 4°C
- Peso Neto: 100-105 gramos (g)
- Unidades por envase: 7 lonchas

Se tomaron y midieron las siguientes muestras, registrándose los siguientes resultados.

MUESTRAS DE PRODUCTO: ROAST BEEF LONCHEADO			
Fecha y hora	Temperatura	Peso	N° de lonchas
1. 04/06: 9:30 am	0,5 °C	102 g	7
2. 04/06: 10:15 am	1,8 °C	105 g	7
3. 04/06: 10:45 am	2,5 °C	108 g	7
4. 04/06: 11:16 am	3,2 °C	101 g	7
5. 04/06: 11:50 am	4,2 °C	104 g	7

- **Prompt:** ¿Cuál de las cinco muestras cumple con todos los requisitos de control de calidad?
- **Opciones:**
 1. **Muestra 1**
 2. **Muestra 2**
 3. Muestra 3
 4. **Muestra 4**
 5. Muestra 5

Prueba de habilidades académicas (PAA):

■ Redacción (Writing):

1. Ítem 1:

• Inciso:

(1) En la zona histórica de Guayama, Ciudad de los Brujos, se encuentra el Teatro Guayama conocido antes como Teatro Calimano, una de las edificaciones más destacadas de este pueblo costero del sur de Puerto Rico. (2) Es una antigua estructura que sobresale entre las abundantes construcciones de influencia criolla y francesa por su estilo Art Deco que rompe con el patrón arquitectónico tradicional del centro urbano. (3) Está situado frente a la plaza pública en la calle Derke, una de las calles del casco urbano del pueblo. (4) En sus paredes, parece haberse detenido el tiempo. (5) Fue construido en el 1938, (6) remodelado en el año 1993 (7) y aún conserva su fachada original.

(8) La estructura externa del teatro es digna de admirarse.

(9) La edificación posee dos niveles de gran elegancia.

(10) Lo primero que sorprende al que llega, visitante isleño o extranjero, es la solidez de sus formas; (11) pero, sobre todo,

destaca el diseño utilizado en su decoración. (12) Líneas

horizontales y verticales en la fachada hacen gala de un

armonioso contraste que recrea la vista e invita a examinar el resto de la estructura con detalle. (13) Se puede

disfrutar de su amplia entrada, de sus estilizadas ventanas geométricamente colocadas y de la amplitud del ventanal grande.

(14) Asimismo, su boletería llama la atención porque conserva su atractiva construcción en madera, material utilizado para decorar en el estilo Art Deco. (15) Es como una invitación a disfrutar de los secretos que guarda el interior del Teatro.

- **Prompt:** ¿Cuál enunciado añade información y enriquece ESTILÍSTICAMENTE la secuencia (1–7) sin negar lo expresado?
- **Opciones:**
 - a) ... una de las edificaciones más destacadas de este pueblo... (1)

- b) Es una antigua estructura... (2)
- c) Está situado frente a la plaza pública... (3)
- d) Fue construido en el 1938,... (5)

2. Ítem 2:

- **Inciso:**

(1) En la zona histórica de Guayama, Ciudad de los Brujos, se encuentra el Teatro Guayama conocido antes como Teatro Calimano, una de las edificaciones más destacadas de este pueblo costero del sur de Puerto Rico. (2) Es una antigua estructura que sobresale entre las abundantes construcciones de influencia criolla y francesa por su estilo Art Deco que rompe con el patrón arquitectónico tradicional del centro urbano. (3) Está situado frente a la plaza pública en la calle Derke, una de las calles del casco urbano del pueblo. (4) En sus paredes, parece haberse detenido el tiempo. (5) Fue construido en el 1938, (6) remodelado en el año 1993 (7) y aún conserva su fachada original.

(8) La estructura externa del teatro es digna de admirarse.

(9) La edificación posee dos niveles de gran elegancia.

(10) Lo primero que sorprende al que llega, visitante isleño o extranjero, es la solidez de sus formas; (11) pero, sobre todo, destaca el diseño utilizado en su decoración. (12) Líneas horizontales y verticales en la fachada hacen gala de un armonioso contraste que recrea la vista e invita a examinar el resto de la estructura con detalle. (13) Se puede disfrutar de su amplia entrada, de sus estilizadas ventanas geométricamente colocadas y de la amplitud del ventanal grande. (14) Asimismo, su boletería llama la atención porque conserva su atractiva construcción en madera, material utilizado para decorar en el estilo Art Deco. (15) Es como una invitación a disfrutar de los secretos que guarda el interior del Teatro.

- **Prompt:** ¿Cuál enunciado añade información y enriquece ESTILÍSTICAMENTE la secuencia (1–7) sin negar lo expresado?

- **Opciones:**

- a) ... Ciudad de los Brujos,... (1)
- b) ... sobresale entre las abundantes construcciones... (2)
- c) ... rompe con el patrón arquitectónico tradicional... (2)
- d) ... parece haberse detenido el tiempo. (4)

■ Entendimiento del inglés (english comprehension):

1. Ítem 1:

- **Inciso:**

Sandra - - - yesterday because she was injured.

- **Opciones:**

- a)* works
- b)* worked
- c)* didn't work
- d)* doesn't work

2. **Ítem 2:**

- **Inciso:**

Figura 33. PAA - Entendimiento del inglés: Pregunta 22

An Excellent Opportunity!

**Part-time
Cat Care Associate**

Our company is expanding and we're looking for the best applicants to join our world-famous crew! Candidates must enjoy animals, love seeing bright eyes, and appreciate hearing meowing voices. Experience is preferred, but we can train the right person.

Applications are being accepted for a limited time.

Please apply at our Catcare Center:
Las Vegas Catcare
1010 Airport Rd.
Las Vegas, NV 89104

- **Opciones:**

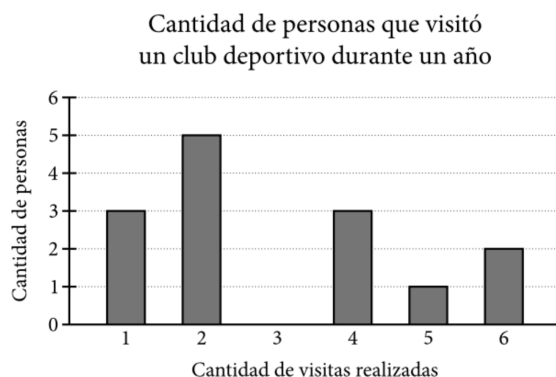
- a)* telephone the company.
- b)* wait for an invitation.
- c)* come to the office.
- d)* email a note.

- **Razonamiento matemático (mathematical reasoning):**

1. **Ítem 1:**

- **Inciso:**

Figura 34. PAA - Razonamiento matemático: Pregunta 17



- **Prompt:** En la gráfica, ¿cuál es la mediana de la cantidad de visitas realizadas al club deportivo?

- **Opciones:**

- a) 2
- b) 2.5
- c) 3
- d) 3.5

2. Ítem 2:

- **Inciso:** ¿Cuál de las siguientes opciones es solución de $4x^2 - 8 = 56$?

- **Opciones:**

- a) 4
- b) 8
- c) 16
- d) 32

11.1.2. Resultados módulo de comunicación

Endpoints de la API

- **/initial-test:** endpoints relacionados con la prueba inicial de evaluación del usuario.

- **GET /initial-test/items:** se encargará de obtener los ítems de la prueba inicial de evaluación del usuario.

- Entrada (Request): Ninguna.
- Salida (Response):

```
Array<{  
  "_id": Id documento en mongo,  
  "id": Id del ítem,  
  "section": Sección del ítem,  
  "type": Tipo de pregunta,  
  "prompt": Pregunta,  
}
```

```

    "options": Lista de opciones de respuesta,
  }>

```

- **POST /initial-test/calculate:** se encargará de obtener las recomendaciones de carreras universitarias utilizando los resultados de la prueba inicial de evaluación del usuario.

- Entrada (Request):

```

{
  "header": {
    "user_id": "1234567890",
  },
  "body": {
    "answers": Array<{
      "item_id": Respuesta del usuario para el ítem,
    }>
  }
}

```

- Salida (Response):

```

{
  "recommendations": Array<{
    "career": Información de la carrera
    recomendada,
    "score": Puntaje de compatibilidad entre las
    características del usuario y la carrera
    recomendada,
    "compatibility_pct": Porcentaje de
    compatibilidad entre las características del
    usuario y la carrera recomendada,
    "why_pos": Explicación positiva de la
    compatibilidad entre las características
    del usuario y la carrera recomendada,
    "why_neg": Explicación negativa de la
    compatibilidad entre las características
    del usuario y la carrera recomendada,
    "enhanced_explanation": Explicación mejorada
    de la compatibilidad entre las características
    del usuario y la carrera recomendada,
  }>,
}

```

- **/guide:** endpoints relacionados con la guía de preguntas del usuario.

- **POST /guide/ask_without_context:** se encargará de responder al usuario sin contexto.

- Entrada (Request):

```
{
  body: {
    "question": Pregunta del usuario,
  },
}
```

- Salida (Response):

```
{
  "success": true,
  "question": Pregunta del usuario,
  "answer": Respuesta del usuario,
}
```

- **POST /guide/ask:** se encargará de responder al usuario con contexto de la base de datos de información académica.

- Entrada (Request):

```
{
  body: {
    "question": Pregunta del usuario,
    "history": Historial de preguntas y respuestas,
  },
}
```

- Salida (Response):

```
{
  "question": Pregunta del usuario,
  "answer": Respuesta del usuario,
  "sources": Fuentes de información consultadas de la
base de datos de información académica para
responder la pregunta.
}
```

- **/generate:** se encargará de generar contenido temático basado en las características del usuario.

- **POST /generate/career-combination/:careerId:** se encargará de obtener las preguntas de la generación de contenido temático del usuario.

- Entrada (Request):

```
{
  params: {
    "career_id": Id de la carrera,
  },
}
```

- Salida (Response):

```
{
  "newCareerId": Id de la nueva carrera,
  "shortQuestion": Pregunta corta de la nueva carrera,
  "completeQuestion": Pregunta completa de la nueva carrera,
}
```

- **POST /generate/mini-inform:** se encargará de obtener las respuestas de la generación de contenido temático del usuario.

- Entrada (Request):

```
{
  body: {
    "question": Pregunta del usuario,
  },
}
```

- Salida (Response):

```
{
  "mini_inform": Mini informe de 100 palabras explicando la
pregunta del usuario,
}
```

- **POST /generate/card:** se encargará de obtener las respuestas de la generación de contenido temático del usuario.

- Entrada (Request):

```
{
  body: {
    "card_type": Tipo de tarjeta (encuesta/what_if),
  },
}
```

- Salida (Response):

```
{
  "card": Tarjeta generada por el modelo de IA, basado en la
base de datos de información académica.
}
```

- **/career:** se encargará de recomendar carreras universitarias basadas en las características del usuario.

- **POST /career/similar/:careerId:** se encargará de obtener las preguntas de la recomendación de carreras universitarias del usuario.

- Entrada (Request):

```
{
  params: {
    "career_id": Id de la carrera,
  },
}
```

```

    }

    ○ Salida (Response):
    {
      "similar_career": Array<{
        "career_id": Id de la carrera,
        "name": Nombre de la carrera,
        "image_url": URL de la imagen de la carrera,
        "description": Descripción de la carrera,
        "duration": Duración de la carrera,
        "employability": Empleabilidad de la carrera,
        "faculty": Facultad de la carrera,
        "similarity_score": Puntaje de similitud entre la
        carrera buscada y la carrera,
      }>,
    }

```

11.1.3. Resultados módulo de guianza vocacional

Preguntas y métricas de la prueba

Pregunta 1: ¿Cuántos años dura Ingeniería en Ciencias de Alimentos?

```

{
  "question": "¿Cuántos años dura Ingeniería en Ciencias de Alimentos?",
  "answer": "¡Hola! La carrera de Ingeniería en Ciencias de Alimentos
  ↪ tiene una duración de 5 años. Considerando tu perfil, donde destacas
  ↪ en áreas como números, resolución de problemas y habilidades
  ↪ investigativas, esta carrera podría ser una excelente opción para
  ↪ ti. ¿Te gustaría saber más sobre las áreas en las que podrías
  ↪ trabajar como ingeniero en alimentos?\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d0ff",
      "collection": "careers",
      "title": "Ingenieria en Ciencias de Alimentos",
      "section": "overview",
      "relevanceScore": 0.787125349
    },
    {
      "id": "68e03f2bae67287b8834d0ff",
      "collection": "careers",
      "title": "Ingenieria en Ciencias de Alimentos",
      "section": "training_area:gerencia_administracion_y_gestion_de_
      ↪ negocios_de_alimentos",
    }
  ]
}

```

```

    "relevanceScore": 0.738691211
  },
  {
    "id": "68e03f2bae67287b8834d0ff",
    "collection": "careers",
    "title": "Ingenieria en Ciencias de Alimentos",
    "section": "training_area:gestion_aseguramiento_y_control_de_ca_
↪ lidad_e_inocuidad",
    "relevanceScore": 0.731910169
  }
]
}

```

Cuadro 20. Resultados de la pregunta 1

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 2: ¿A qué facultad pertenece Ingeniería Mecatrónica?

```

{
  "question": "¿A qué facultad pertenece Ingeniería Mecatrónica?",
  "answer": "¡Hola! Ingeniería Mecatrónica pertenece a la Facultad de
↪ Ingeniería. Esta carrera combina diseño electrónico y mecánico para
↪ crear sistemas mecatrónicos, con un enfoque en robótica y
↪ automatización. Considerando tus altos puntajes en resolución de
↪ problemas y habilidades numéricas, además de tu interés en todas las
↪ áreas RIASCE, esta carrera podría ser una excelente opción para
↪ ti.",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d109",
      "collection": "careers",
      "title": "Ingenieria Mecatronica",
      "section": "overview",
      "relevanceScore": 0.750933
    },
    {
      "id": "68e03f2bae67287b8834d109",
      "collection": "careers",
      "title": "Ingenieria Mecatronica",
      "section": "training_area:maquinaria_inteligente",
      "relevanceScore": 0.74513334
    }
  ]
}

```

```

    },
    {
      "id": "68e03f2bae67287b8834d109",
      "collection": "careers",
      "title": "Ingenieria Mecatronica",
      "section": "training_area:automatizacion",
      "relevanceScore": 0.741915882
    }
  ]
}

```

Cuadro 21. Resultados de la pregunta 2

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 3: ¿Cuál es la empleabilidad de la Ingeniería Mecánica?

```

{
  "question": "¿Cuál es la empleabilidad de la Ingeniería Mecánica?",
  "answer": "¡Hola! La empleabilidad de la Ingeniería Mecánica es muy
    ↪ alta. Esto significa que hay muchas oportunidades de trabajo
    ↪ disponibles para los graduados en este campo. Además, tus resultados
    ↪ en las pruebas indican una alta capacidad para la resolución de
    ↪ problemas y un buen nivel en matemáticas, lo que te convertiría en
    ↪ un candidato muy competitivo en este campo. ¿Hay algo más en lo que
    ↪ te pueda ayudar?\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d10a",
      "collection": "careers",
      "title": "Ingenieria Mecanica",
      "section": "overview",
      "relevanceScore": 0.752710521
    },
    {
      "id": "68e03f2bae67287b8834d10c",
      "collection": "careers",
      "title": "Ingenieria Mecanica Industrial",
      "section": "overview",
      "relevanceScore": 0.732152343
    }
  ]
}

```

```

        "id": "68e03f2bae67287b8834d109",
        "collection": "careers",
        "title": "Ingenieria Mecatronica",
        "section": "overview",
        "relevanceScore": 0.725645483
    }
}

```

Cuadro 22. Resultados de la pregunta 3

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 4: ¿Cómo es el perfil del egresado de la carrera de Ingeniería Civil Arquitectónica?

```

{
  "question": "¿Cómo es el perfil del egresado de la carrera de Ingeniería  

  ↳ Civil Arquitectónica?",
  "answer": "¡Hola! El perfil del egresado de Ingeniería Civil  

  ↳ Arquitectónica se especializa en el análisis y diseño arquitectónico  

  ↳ y estructural. También está capacitado para diseñar instalaciones  

  ↳ hidráulicas y electromecánicas, así como para encargarse de la  

  ↳ construcción y el mantenimiento de edificaciones. Todo esto se  

  ↳ realiza con un fuerte énfasis en el uso y la conservación de los  

  ↳ recursos naturales.\n\nConsiderando tus resultados en las pruebas,  

  ↳ tu alta puntuación en razonamiento numérico y resolución de  

  ↳ problemas (piaac), junto con un interés equilibrado en todas las  

  ↳ áreas RISEC, sugiere una buena compatibilidad con esta carrera.\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d101",
      "collection": "careers",
      "title": "Ingenieria Civil Arquitectonica",
      "section": "graduate_profile",
      "relevanceScore": 0.835799456
    },
    {
      "id": "68e03f2bae67287b8834d107",
      "collection": "careers",
      "title": "Ingenieria Civil",
      "section": "graduate_profile",
    }
  ]
}

```

```

    "relevanceScore": 0.804933846
  },
  {
    "id": "68e03f2bae67287b8834d10e",
    "collection": "careers",
    "title": "Licenciatura en Arquitectura",
    "section": "graduate_profile",
    "relevanceScore": 0.786942899
  }
]
}

```

Cuadro 23. Resultados de la pregunta 4

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 5: ¿La carrera de Ingeniería Electrónica lleva la clase de Cálculo 1? ¿Si sí, en que año y semestre la lleva?

```

{
  "question": "¿La carrera de Ingeniería Electrónica lleva la clase de  

  ↳ Cálculo 1? ¿Si sí, en que año y semestre la lleva?",
  "answer": "¡Hola! Sí, la carrera de Ingeniería Electrónica incluye  

  ↳ Cálculo 1 en su plan de estudios. La llevas en el primer año,  

  ↳ durante el segundo semestre. Con tus altas habilidades en  

  ↳ matemáticas (num=1 en piaac y paa), ¡seguro te irá muy bien en esa  

  ↳ clase!",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d105",
      "collection": "careers",
      "title": "Ingenieria Electronica",
      "section": "study_plan:y3:primer_semestre",
      "relevanceScore": 0.709406912
    },
    {
      "id": "68e03f2bae67287b8834d105",
      "collection": "careers",
      "title": "Ingenieria Electronica",
      "section": "study_plan:y1:segundo_semestre",
      "relevanceScore": 0.708833337
    }
  ],
}

```

```

    {
      "id": "68e03f2bae67287b8834d105",
      "collection": "careers",
      "title": "Ingenieria Electronica",
      "section": "study_plan:y1:primer_semestre",
      "relevanceScore": 0.705986559
    }
  ]
}

```

Cuadro 24. Resultados de la pregunta 5

Métrica	Valor
Exactitud	5
Relevancia	4.67
Claridad	4.67
Puntaje promedio	4.78
Fidelidad	95.56 %

Pregunta 6: ¿La carrera de Sistemas Computacionales lleva muchos cursos de matemática? ¿Si sí, podrías proporcionarme algunos ejemplos?

```

{
  "question": "¿La carrera de Sistemas Computacionales lleva muchos cursos  

  ↪ de matemática? ¿Si sí, podrías proporcionarme algunos ejemplos?",
  "answer": "¡Hola! Veo que te interesa la Licenciatura en Sistemas de  

  ↪ Información Computacional. ¡Qué bien!\n\nEfectivamente, esta carrera  

  ↪ sí incluye varios cursos de matemáticas. Algunos ejemplos son  

  ↪ \"Cálculo para los Negocios 1\", \"Matemática Financiera\",  

  ↪ \"Herramientas de Estadística 1\", \"Cálculo para los Negocios 2\" y  

  ↪ \"Álgebra Matricial\". Estos cursos te proporcionarán las bases  

  ↪ necesarias para resolver problemas complejos en el campo de la  

  ↪ informática.\n\nConsiderando tus resultados en las pruebas, tu alta  

  ↪ puntuación en razonamiento numérico (num: 1) sugiere que tienes una  

  ↪ buena capacidad para las matemáticas, lo cual te será muy útil en  

  ↪ esta carrera. ¡Mucho éxito!\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d104",
      "collection": "careers",
      "title": "Licenciatura en Sistemas de Informacion  

      ↪ Computacional",
      "section": "study_plan:y2:segundo_semestre",
      "relevanceScore": 0.691207826
    },
    {
      "id": "68e03f2bae67287b8834d104",

```

```

    "collection": "careers",
    "title": "Licenciatura en Sistemas de Informacion
    ↪ Computacional",
    "section": "study_plan:y1:segundo_semestre",
    "relevanceScore": 0.690974236
  },
  {
    "id": "68e03f2bae67287b8834d11a",
    "collection": "careers",
    "title": "Licenciatura en Matematica Aplicada",
    "section": "study_plan:y2:segundo_semestre",
    "relevanceScore": 0.690855443
  }
]
}

```

Cuadro 25. Resultados de la pregunta 6

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 7: ¿Cuántos cursos se lleva en el tercer año de la carrera de Química Farmacéutica?

```

{
  "question": "¿Cuántos cursos se lleva en el tercer año de la carrera de
  ↪ Química Farmacéutica?",
  "answer": "¡Hola! En el tercer año de la Licenciatura en Química
  ↪ Farmacéutica, se llevan 6 cursos por semestre. En el primer
  ↪ semestre, verás materias como Química Orgánica 3 y Bioquímica de
  ↪ Macromoléculas. Y si te interesa, en el segundo semestre estudiarás
  ↪ Fisicoquímica 1 y Bioquímica del Metabolismo Celular. ¡Es un año muy
  ↪ completo!\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d116",
      "collection": "careers",
      "title": "Licenciatura en Quimica Farmaceutica",
      "section": "study_plan:y3:primer_semestre",
      "relevanceScore": 0.750375807
    },
    {
      "id": "68e03f2bae67287b8834d116",

```

```

    "collection": "careers",
    "title": "Licenciatura en Quimica Farmaceutica",
    "section": "study_plan:y3:segundo_semestre",
    "relevanceScore": 0.749754786
  },
  {
    "id": "68e03f2bae67287b8834d119",
    "collection": "careers",
    "title": "Licenciatura en Quimica",
    "section": "study_plan:y3:primer_semestre",
    "relevanceScore": 0.729482472
  }
]
}

```

Cuadro 26. Resultados de la pregunta 7

Métrica	Valor
Exactitud	5
Relevancia	4.67
Claridad	5
Puntaje promedio	4.89
Fidelidad	97.78 %

Pregunta 8: ¿Cuántas físicas se lleva en la carrera de Ingeniería electrónica?

```

{
  "question": "¿Cuántas físicas se lleva en la carrera de Ingeniería  

  ↳ electrónica?",
  "answer": "En el plan de estudios de Ingeniería Electrónica, se cursan  

  ↳ dos cursos de Física: Física 1 en el segundo semestre del primer año  

  ↳ y Física 2 en el primer semestre del segundo año. Tu puntaje alto en  

  ↳ resolución de problemas (ps: 1) sugiere que podrías disfrutar estos  

  ↳ cursos, ya que requieren habilidades analíticas y de resolución de  

  ↳ problemas.",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d105",
      "collection": "careers",
      "title": "Ingenieria Electronica",
      "section": "overview",
      "relevanceScore": 0.729266882
    },
    {
      "id": "68e03f2bae67287b8834d105",
      "collection": "careers",
      "title": "Ingenieria Electronica",

```

```

    "section": "training_area:investigacion_de_tecnologia_electroni_
    ↪ ca_y_robotica",
    "relevanceScore": 0.707209527
  },
  {
    "id": "68e03f2bae67287b8834d105",
    "collection": "careers",
    "title": "Ingenieria Electronica",
    "section": "study_plan:y3:primer_semestre",
    "relevanceScore": 0.706497669
  }
]
}

```

Cuadro 27. Resultados de la pregunta 8

Métrica	Valor
Exactitud	4.67
Relevancia	4.67
Claridad	5
Puntaje promedio	4.78
Fidelidad	95.56 %

Pregunta 9: ¿Podrías mencionarme algunas de las clases más complicadas de la carrera de ciencias de la computación?

```

{
  "question": "¿Podrías mencionarme algunas de las clases más complicadas
  ↪ de la carrera de ciencias de la computación?",
  "answer": "¡Claro! Viendo el plan de estudios de Ingeniería en Ciencia
  ↪ de la Computación y Tecnologías de la Información, algunas de las
  ↪ clases que podrían ser más desafiantes son \"Análisis y Diseño de
  ↪ Algoritmos\", \"Teoría de la Computación\" y \"Construcción de
  ↪ Compiladores\". Estas materias requieren un pensamiento lógico y
  ↪ matemático bastante desarrollado, lo cual se alinea bien con tu alta
  ↪ puntuación en habilidades numéricas y de resolución de problemas.
  ↪ ¿Te gustaría saber algo más sobre alguna de estas clases en
  ↪ particular?\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d103",
      "collection": "careers",
      "title": "Ingenieria en Ciencia de la Computacion y Tecnologias
      ↪ de la Informacion",
      "section": "study_plan:y4:segundo_semestre",
      "relevanceScore": 0.690342069
    }
  ],
}

```

```

{
  "id": "68e03f2bae67287b8834d103",
  "collection": "careers",
  "title": "Ingenieria en Ciencia de la Computacion y Tecnologias
↪ de la Informacion",
  "section": "competency:2",
  "relevanceScore": 0.679784775
},
{
  "id": "68e03f2bae67287b8834d103",
  "collection": "careers",
  "title": "Ingenieria en Ciencia de la Computacion y Tecnologias
↪ de la Informacion",
  "section": "development_area:algoritmos_y_teoría_computacional",
  "relevanceScore": 0.677139163
}
]
}

```

Cuadro 28. Resultados de la pregunta 9

Métrica	Valor
Exactitud	5
Relevancia	4.67
Claridad	5
Puntaje promedio	4.89
Fidelidad	97.78 %

Pregunta 10: ¿Qué clases, pertenecientes a mis carreras recomendadas, podrían gustarme más?

```

{
  "question": "¿Qué clases, pertenecientes a mis carreras recomendadas,
↪ podrían gustarme más?",
  "answer": "Considerando tus carreras recomendadas y tus resultados, es
↪ probable que disfrutes las clases de química orgánica, biología
↪ celular y bioquímica en la Licenciatura en Química Farmacéutica.
↪ También podrías encontrar interesantes las clases de diseño de
↪ ingeniería, termodinámica y mecánica de fluidos en Ingeniería
↪ Mecatrónica. Tus altas puntuaciones en razonamiento numérico y
↪ resolución de problemas sugieren que te destacarás en estas
↪ áreas.\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d116",
      "collection": "careers",
      "title": "Licenciatura en Química Farmacéutica",

```

```

    "section": "development_area:diferentes_campos",
    "relevanceScore": 0.67304796
  },
  {
    "id": "68e03f2bae67287b8834d113",
    "collection": "careers",
    "title": "Licenciatura en Psicología",
    "section": "tags",
    "relevanceScore": 0.669339
  },
  {
    "id": "68e03f2bae67287b8834d10b",
    "collection": "careers",
    "title": "Licenciatura en Comunicacion Estrategica",
    "section": "tags",
    "relevanceScore": 0.665144503
  }
]
}

```

Cuadro 29. Resultados de la pregunta 10

Métrica	Valor
Exactitud	5
Relevancia	4.67
Claridad	5
Puntaje promedio	4.89
Fidelidad	97.78 %

Pregunta 11: ¿En qué tipo de proyectos o roles trabajan los graduados en Ciencias de Alimentos?

```

{
  "question": "¿En qué tipo de proyectos o roles trabajan los graduados en  

  ↳ Ciencias de Alimentos?",
  "answer": "¡Hola! Los graduados en Ingeniería en Ciencias de Alimentos  

  ↳ tienen un campo laboral muy amplio. Pueden participar en el  

  ↳ desarrollo de nuevos productos alimenticios, diseñar procesos  

  ↳ eficientes para la producción de alimentos, gestionar proyectos de  

  ↳ investigación y asegurar la calidad e inocuidad en la industria  

  ↳ alimentaria.\n\nConsiderando tus resultados en los tests,  

  ↳ especialmente tu alta capacidad de resolución de problemas y tu  

  ↳ interés en áreas realistas e investigativas, esta carrera podría ser  

  ↳ muy compatible contigo. ¿Te gustaría saber más sobre alguna de estas  

  ↳ áreas en particular?\n",
  "sources": [
    {

```

```

      "id": "68e03f2bae67287b8834d0ff",
      "collection": "careers",
      "title": "Ingenieria en Ciencias de Alimentos",
      "section": "graduate_profile",
      "relevanceScore": 0.744754136
    },
    {
      "id": "68e03f2bae67287b8834d0ff",
      "collection": "careers",
      "title": "Ingenieria en Ciencias de Alimentos",
      "section":
        ↪ "development_area:investigacion_y_gestion_de_proyectos",
      "relevanceScore": 0.741140783
    },
    {
      "id": "68e03f2bae67287b8834d11b",
      "collection": "careers",
      "title": "Licenciatura en Nutricion",
      "section":
        ↪ "development_area:ciencias_y_tecnologia_de_alimentos",
      "relevanceScore": 0.732749522
    }
  ]
}

```

Cuadro 30. Resultados de la pregunta 11

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 12: ¿Qué áreas se desarrollan dentro del diseño de productos?

```

{
  "question": "¿Qué áreas se desarrollan dentro del diseño de productos?",
  "answer": "¡Hola! En la Licenciatura en Diseño de Producto e Innovación,
    ↪ te formarás en áreas como el pensamiento de diseño para entender a
    ↪ los consumidores, la creación de soluciones integrales e
    ↪ innovadoras, la identificación de oportunidades emergentes y la
    ↪ humanización de productos para un desarrollo sostenible.
    ↪ ¡Considerando tu alta puntuación en áreas como resolución de
    ↪ problemas y tu perfil RIASEC equilibrado, esta carrera podría ser
    ↪ muy compatible contigo! ¿Te gustaría saber más sobre alguna de estas
    ↪ áreas en particular?\n",

```

```

"sources": [
  {
    "id": "68e03f2bae67287b8834d11d",
    "collection": "careers",
    "title": "Licenciatura en Diseño de Producto e Innovacion",
    "section": "development_area:humanizacion_de_productos",
    "relevanceScore": 0.757716596
  },
  {
    "id": "68e03f2bae67287b8834d11d",
    "collection": "careers",
    "title": "Licenciatura en Diseño de Producto e Innovacion",
    "section": "development_area:pensamiento_de_diseno",
    "relevanceScore": 0.734693944
  },
  {
    "id": "68e03f2bae67287b8834d11d",
    "collection": "careers",
    "title": "Licenciatura en Diseño de Producto e Innovacion",
    "section": "training_area:colaboracion_tecnologica",
    "relevanceScore": 0.733671606
  }
]
}

```

Cuadro 31. Resultados de la pregunta 12

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 13: ¿Qué competencias se desarrollan al estudiar psicología?

```

{
  "question": "¿Qué competencias se desarrollan al estudiar psicología?",

```

```

"answer": "¡Hola! Con gusto te cuento sobre las competencias que se
↳ desarrollan al estudiar la Licenciatura en Psicología.\n\nSegún la
↳ información que tengo, algunas de las competencias que desarrollarás
↳ son:\n\n* Implementar procesos de evaluación utilizando
↳ herramientas basadas en evidencia científica.\n* Diseñar e
↳ implementar intervenciones para promover el bienestar y la salud
↳ mental.\n* Crear e implementar intervenciones psicoeducativas para
↳ promover competencias psicológicas.\n* Implementar procesos de
↳ análisis y diagnóstico basados en evaluaciones integrales.\n*
↳ Diseñar e implementar proyectos para promover el éxito
↳ organizacional y el bienestar.\n* Diseñar e implementar
↳ investigaciones rigurosas para contribuir al avance
↳ científico.\n\nConsiderando tus resultados en las pruebas, tu alta
↳ puntuación en las áreas social y de apertura (S y O en RIASEC y BFI,
↳ respectivamente) sugiere una buena compatibilidad con la
↳ Licenciatura en Psicología, ya que estas áreas se relacionan con la
↳ empatía, la comprensión interpersonal y el interés por el bienestar
↳ de los demás, habilidades cruciales en esta disciplina. ¿Te gustaría
↳ saber algo más sobre esta carrera?\n",
"sources": [
  {
    "id": "68e03f2bae67287b8834d113",
    "collection": "careers",
    "title": "Licenciatura en Psicologia",
    "section": "competency:4",
    "relevanceScore": 0.733297
  },
  {
    "id": "68e03f2bae67287b8834d113",
    "collection": "careers",
    "title": "Licenciatura en Psicologia",
    "section": "competency:3",
    "relevanceScore": 0.729401648
  },
  {
    "id": "68e03f2bae67287b8834d113",
    "collection": "careers",
    "title": "Licenciatura en Psicologia",
    "section": "competency:6",
    "relevanceScore": 0.729168773
  }
]
}

```

Pregunta 14: ¿Cuál es el salario de un Ingeniero Mecánico?

```
{
```

Cuadro 32. Resultados de la pregunta 13

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

```

"question": "¿Cuál es el salario de un Ingeniero Mecánico?",
"answer": "¡Hola! Un ingeniero mecánico puede esperar un salario que
↪ oscila entre $60,780 y $155,590 USD. Considerando tus altas
↪ puntuaciones en resolución de problemas y habilidades numéricas,
↪ ¡esta podría ser una excelente opción para ti! ¿Te gustaría saber
↪ algo más sobre esta carrera o alguna otra opción?\n",
"sources": [
  {
    "id": "68e03f2bae67287b8834d10a",
    "collection": "careers",
    "title": "Ingenieria Mecanica",
    "section": "salary",
    "relevanceScore": 0.781853378
  },
  {
    "id": "68e03f2bae67287b8834d10c",
    "collection": "careers",
    "title": "Ingenieria Mecanica Industrial",
    "section": "salary",
    "relevanceScore": 0.76999265
  },
  {
    "id": "68e03f2bae67287b8834d109",
    "collection": "careers",
    "title": "Ingenieria Mecatronica",
    "section": "salary",
    "relevanceScore": 0.752856314
  }
]
}

```

Pregunta 15: ¿Cuál es la diferencia entre las carreras de química y química farmacéutica?

```

{
  "question": "¿Cuál es la diferencia entre las carreras de química y
↪ química farmacéutica?",

```

Cuadro 33. Resultados de la pregunta 14

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

```

"answer": "¡Hola! La principal diferencia entre Química y Química
↪ Farmacéutica radica en su enfoque. Mientras que Química abarca un
↪ estudio amplio de la materia y sus propiedades, Química Farmacéutica
↪ se centra en el diseño, desarrollo, producción y control de calidad
↪ de medicamentos y cosméticos.\n\nConsiderando tus resultados en los
↪ tests, especialmente tu alta puntuación en razonamiento numérico y
↪ resolución de problemas, junto con tu interés en áreas
↪ investigativas y tu nivel de conciencia, Química Farmacéutica podría
↪ ser una excelente opción para ti, ya que combina la rigurosidad de
↪ la química con aplicaciones prácticas en el campo de la salud.",
"sources": [
  {
    "id": "68e03f2bae67287b8834d116",
    "collection": "careers",
    "title": "Licenciatura en Quimica Farmaceutica",
    "section": "development_area:diferentes_campos",
    "relevanceScore": 0.749303281
  },
  {
    "id": "68e03f2bae67287b8834d116",
    "collection": "careers",
    "title": "Licenciatura en Quimica Farmaceutica",
    "section": "training_area:ciencias_farmaceuticas",
    "relevanceScore": 0.735916197
  },
  {
    "id": "68e03f2bae67287b8834d116",
    "collection": "careers",
    "title": "Licenciatura en Quimica Farmaceutica",
    "section": "overview",
    "relevanceScore": 0.734215617
  }
]
}

```

Pregunta 16: ¿Los ingenieros electrónicos tienden a ganar más que los ingenieros mecatrónicos?

Cuadro 34. Resultados de la pregunta 15

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

```
{
  "question": "¿Los ingenieros electrónicos tienden a ganar más que los
    ↪ ingenieros mecatrónicos?",
  "answer": "¡Hola! Basado en la información que tengo, los ingenieros
    ↪ electrónicos tienden a tener un rango de salario ligeramente más
    ↪ alto que los ingenieros mecatrónicos. El salario mínimo para
    ↪ ingenieros electrónicos es de $71,830 USD y el máximo es de $182,310
    ↪ USD, mientras que para los mecatrónicos es de $60,780 USD a $155,590
    ↪ USD.\n\nConsiderando tus resultados, tienes una alta compatibilidad
    ↪ con ambas ingenierías (88.2 para mecatrónica). ¿Hay algo en
    ↪ particular que te atrae de alguna de estas dos opciones? ¡Estaré
    ↪ aquí para ayudarte a explorar tus intereses!\n",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d109",
      "collection": "careers",
      "title": "Ingenieria Mecatronica",
      "section": "salary",
      "relevanceScore": 0.71975565
    },
    {
      "id": "68e03f2bae67287b8834d105",
      "collection": "careers",
      "title": "Ingenieria Electronica",
      "section": "salary",
      "relevanceScore": 0.706143796
    },
    {
      "id": "68e03f2bae67287b8834d109",
      "collection": "careers",
      "title": "Ingenieria Mecatronica",
      "section": "development_area:diseno_electronico",
      "relevanceScore": 0.694663644
    }
  ]
}
```

Pregunta 17: ¿Por qué crees que se debe? (Continuación de Pregunta 16)

Cuadro 35. Resultados de la pregunta 16

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

```
{
  "question": "¿Por qué crees que se debe?",
  "answer": "¡Claro! La diferencia salarial entre Ingeniería Electrónica y Mecatrónica p",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d119",
      "collection": "careers",
      "title": "Licenciatura en Quimica",
      "section": "development_area:sintesis_quimica",
      "relevanceScore": 0.603886366
    },
    {
      "id": "68e03f2bae67287b8834d10b",
      "collection": "careers",
      "title": "Licenciatura en Comunicacion Estrategica",
      "section": "training_area:divulgacion_de_politicas_publicas",
      "relevanceScore": 0.601895809
    },
    {
      "id": "68e03f2bae67287b8834d118",
      "collection": "careers",
      "title": "Licenciatura en Antropologia",
      "section": "competency:1",
      "relevanceScore": 0.596568644
    },
    {
      "id": "68e03f2bae67287b8834d109",
      "collection": "careers",
      "title": "Ingenieria Mecatronica",
      "section": "salary",
      "relevanceScore": 0.739160597
    },
    {
      "id": "68e03f2bae67287b8834d105",
      "collection": "careers",
      "title": "Ingenieria Electronica",
      "section": "salary",
      "relevanceScore": 0.727926731
    }
  ]
}
```

```

    }
  ]
}

```

Cuadro 36. Resultados de la pregunta 17

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

Pregunta 18: ¿Cuál es la diferencia entre biología y biomédica?

```

{
  "question": "¿Cuál es la diferencia entre biología y biomédica?",
  "answer": "¡Hola! La Biología es el estudio de la vida en todas sus formas, desde las moléculas hasta los ecosistemas.",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d108",
      "collection": "careers",
      "title": "Licenciatura en Biología",
      "section": "competency:4",
      "relevanceScore": 0.705091238
    },
    {
      "id": "68e03f2bae67287b8834d108",
      "collection": "careers",
      "title": "Licenciatura en Biología",
      "section": "overview",
      "relevanceScore": 0.696818233
    },
    {
      "id": "68e03f2bae67287b8834d108",
      "collection": "careers",
      "title": "Licenciatura en Biología",
      "section": "development_area:investigacion_biologica",
      "relevanceScore": 0.694085181
    }
  ]
}

```

Pregunta 19: ¿Según tus bases de datos, quien gana más un biólogo o un biomédico?

```

{

```

Cuadro 37. Resultados de la pregunta 18

Métrica	Valor
Exactitud	5
Relevancia	5
Claridad	5
Puntaje promedio	5
Fidelidad	100 %

```

"question": "¿Según tus bases de datos, quien gana más un biologo o un biomédico?",
"answer": "¡Buena pregunta! Según mis datos, un ingeniero biomédico tiende a ganar más",
"sources": [
  {
    "id": "68e03f2bae67287b8834d108",
    "collection": "careers",
    "title": "Licenciatura en Biologia",
    "section": "salary",
    "relevanceScore": 0.707954049
  },
  {
    "id": "68e03f2bae67287b8834d115",
    "collection": "careers",
    "title": "Licenciatura en Bioquimica y Microbiologia",
    "section": "salary",
    "relevanceScore": 0.705010533
  },
  {
    "id": "68e03f2bae67287b8834d0fe",
    "collection": "careers",
    "title": "Ingenieria Biomedica",
    "section": "salary",
    "relevanceScore": 0.702586
  }
]
}

```

Cuadro 38. Resultados de la pregunta 19

Métrica	Valor
Exactitud	4.33
Relevancia	4.67
Claridad	4.67
Puntaje promedio	4.56
Fidelidad	91.11 %

Pregunta 20: Según tus datos, ¿Cuál de los dos posee mejor empleabilidad?

```

{
  "question": "Según tus datos, ¿Cuál de los dos posee mejor empleabilidad?",
  "answer": "¡Claro! Según mis datos, Ingeniería Biomédica tiene una empleabilidad muy alta, m",
  "sources": [
    {
      "id": "68e03f2bae67287b8834d114",
      "collection": "careers",
      "title": "Licenciatura en Biotecnología Molecular",
      "section": "development_area:analisis_de_datos",
      "relevanceScore": 0.642496705
    },
    {
      "id": "68e03f2bae67287b8834d104",
      "collection": "careers",
      "title": "Licenciatura en Sistemas de Informacion Computacional",
      "section": "overview",
      "relevanceScore": 0.641190112
    },
    {
      "id": "68e03f2bae67287b8834d10d",
      "collection": "careers",
      "title": "Licenciatura en Administracion de Empresas",
      "section": "tags",
      "relevanceScore": 0.639671266
    },
    {
      "id": "68e03f2bae67287b8834d0fe",
      "collection": "careers",
      "title": "Ingenieria Biomedica",
      "section": "tags",
      "relevanceScore": 0.720697582
    },
    {
      "id": "68e03f2bae67287b8834d108",
      "collection": "careers",
      "title": "Licenciatura en Biologia",
      "section": "overview",
      "relevanceScore": 0.716595948
    }
  ]
}

```

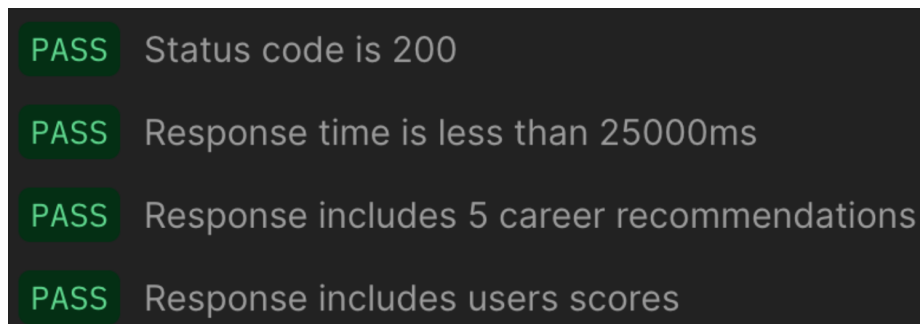
Cuadro 39. Resultados de la pregunta 20

Métrica	Valor
Exactitud	4.33
Relevancia	4.00
Claridad	3.67
Puntaje promedio	4.00
Fidelidad	80.00 %

Figura 35. Resultados de la ejecución de la prueba del objetivo de la guianza

	Escala 1 - 5																		
	Juiz 1			Juiz 2			Juiz 3			Juiz 4									
	Exactitud factual	Relevancia semántica	Claridad y completitud	Exactitud factual	Relevancia semántica	Claridad y completitud	Exactitud factual	Relevancia semántica	Claridad y completitud	Exactitud factual	Relevancia semántica	Claridad y completitud	Total	Fidelidad					
Pregunta 1	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 2	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 3	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 4	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 5	5	4	5	5	5	4	5	5	4	5	5.00	4.67	4.67	95					
Pregunta 6	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 7	5	5	5	5	4	5	5	5	5	5	5.00	4.67	5.00	97.78					
Pregunta 8	5	4	5	5	5	5	4	5	5	5	4.67	4.67	5.00	95.56					
Pregunta 9	5	5	5	5	4	5	5	5	5	5	5.00	4.67	5.00	97.78					
Pregunta 10	5	5	5	5	5	4	5	5	5	5	5.00	4.67	5.00	97.78					
Pregunta 11	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 12	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 13	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 14	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 15	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 16	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 17	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 18	5	5	5	5	5	5	5	5	5	5	5.00	5.00	5.00	100					
Pregunta 19	4	5	5	4	5	4	5	4	5	4	4.33	4.67	4.67	91.11					
Pregunta 20	4	4	3	5	5	3	4	3	5	4	4.33	4.00	3.67	80					
Total	4.9	4.85	4.9	4.95	4.85	4.9	4.9	4.85	4.9	4.85	5								
Fidelidad	98	97	98	99	97	96	98	97	100										
Exactitud Factual	4.92			4.85			4.9			4.9									

Figura 37. Resultados de la ejecución de la prueba del objetivo de la recomendación



Nota. Elaboración propia.

Figura 38. Resultados de la ejecución de la prueba del objetivo de la recomendación

Source	Environment	Iterations	Duration	All tests	Avg. Resp. Time
Runner	Testing	1	15m 30s	80	11507 ms

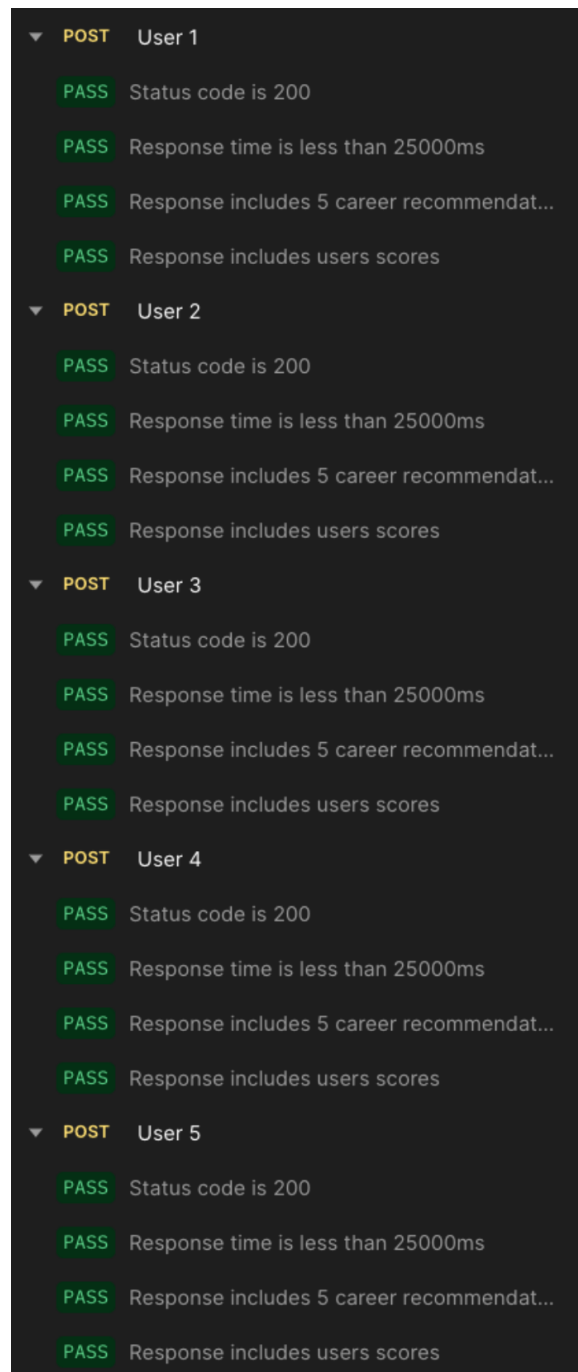
Nota. Elaboración propia.

Figura 39. Resultados de la ejecución de la prueba del objetivo de la recomendación

▶	POST	User 1	4 0	
▶	POST	User 2	4 0	
▶	POST	User 3	4 0	
▶	POST	User 4	4 0	
▶	POST	User 5	4 0	
▶	POST	User 6	4 0	
▶	POST	User 7	4 0	
▶	POST	User 8	4 0	
▶	POST	User 9	4 0	
▶	POST	User 10	4 0	
▶	POST	User 11	4 0	
▶	POST	User 12	4 0	
▶	POST	User 13	4 0	
▶	POST	User 14	4 0	
▶	POST	User 15	4 0	
▶	POST	User 16	4 0	
▶	POST	User 17	4 0	
▶	POST	User 18	4 0	
▶	POST	User 19	4 0	
▶	POST	User 20	4 0	

Nota. Elaboración propia.

Figura 40. Resultados de la ejecución de la prueba del objetivo de la recomendación



▼	POST	User 1
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 2
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 3
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 4
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 5
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores

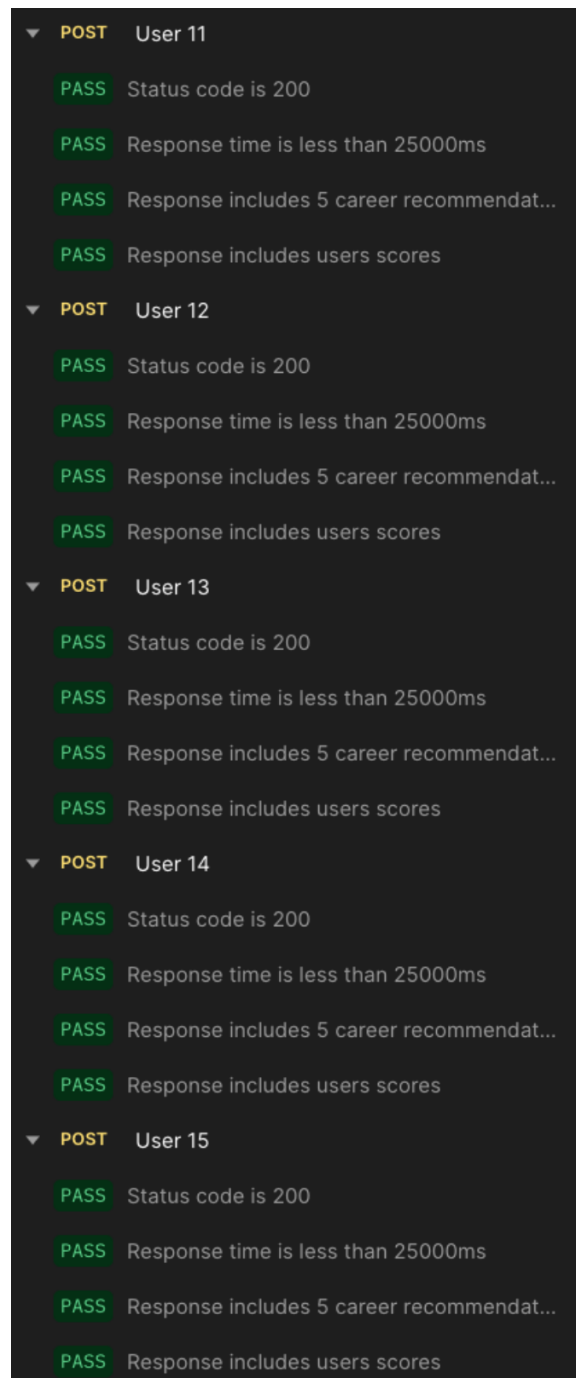
Nota. Elaboración propia.

Figura 41. Resultados de la ejecución de la prueba del objetivo de la recomendación

▼	POST	User 6
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 7
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 8
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 9
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 10
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores

Nota. Elaboración propia.

Figura 42. Resultados de la ejecución de la prueba del objetivo de la recomendación



▼	POST	User 11
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 12
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 13
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 14
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores
▼	POST	User 15
	PASS	Status code is 200
	PASS	Response time is less than 25000ms
	PASS	Response includes 5 career recommendat...
	PASS	Response includes users scores

Nota. Elaboración propia.

Figura 43. Resultados de la ejecución de la prueba del objetivo de la recomendación

▼	POST	User 16
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 17
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 18
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 19
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores
▼	POST	User 20
PASS		Status code is 200
PASS		Response time is less than 25000ms
PASS		Response includes 5 career recommendat...
PASS		Response includes users scores

Nota. Elaboración propia.

Cuadro 40. Ejemplo de datos de prueba para el objetivo de la recomendación

```

{
  // Género de la persona
  "metadata_gender": "mujer",

  // Resultados de la prueba
  ↪ Big Five Personality
  ↪ (BFI-15)
  "bfi_1": "s3",
  "bfi_2": "s1",
  "bfi_3": "s2",
  "bfi_4": "s6",
  "bfi_5": "s4",
  "bfi_6": "s5",
  "bfi_7": "s4",
  "bfi_8": "s5",
  "bfi_9": "s3",
  "bfi_10": "s4",
  "bfi_11": "s6",
  "bfi_12": "s3",
  "bfi_13": "s6",
  "bfi_14": "s5",
  "bfi_15": "s4",

  // Resultados de la prueba
  ↪ RIASEC
  "riasec_1": "r1",
  "riasec_2": "r2",
  "riasec_3": "r4",
  "riasec_4": "r2",
  "riasec_5": "r1",
  "riasec_6": "r1",
  "riasec_7": "r3",
  "riasec_8": "r5",
  "riasec_9": "r1",
  "riasec_10": "r4",
  "riasec_11": "r5",
  "riasec_12": "r3",
}

{
  // Resultados de la prueba
  ↪ Grits
  "grit_1": "g3",
  "grit_2": "g1",
  "grit_3": "g2",
  "grit_4": "g2",
  "grit_5": "g1",
  "grit_6": "g2",
  "grit_7": "g2",
  "grit_8": "g1",

  // Resultados de la prueba
  ↪ PAA
  "paa_redaccion_1": "r1_1",
  "paa_redaccion_2": "r2_4",
  "paa_english_1": "w3",
  "paa_english_2": "e2_3",
  "paa_math_1": "m1_2",
  "paa_math_2": "m2_4",

  // Resultados de la prueba
  ↪ PIAAC
  "piaac_numerica_1": "80",
  "piaac_numerica_2":
    ↪ "t_-16_7",
  "piaac_problem_solving":
    ↪ "ps_2",
  "piaac_reading_1": "09:00",
  "piaac_reading_2": {
    "r1": "pan",
    "r2": "ambos",
    "r3": "ambos"
  }
}

```

Nota. Elaboración propia.

11.1.5. Resultados módulo de comunicación

Figura 44. Tareas realizadas para la prueba del módulo de comunicación

Modulo	Tarea
Autenticación	Crear una cuenta con correo
Autenticación	Cerrar y volver a iniciar sesión
Quiz	Completar todas las preguntas
Quiz	Leer resultados
Exploración	Deslizar 5 tarjetas
Exploración	Guardar una tarjeta
Exploración	Abrir y cerrar una tarjeta
Match	Revisar Top 5 carreras
Match	Abrir una carrera del Top
Carreras	Ver pensum interactivo
Carreras	Usar botón "¿Y si combinas con...?"
Carreras	Filtrar carreras de Facultad de Ingeniería
Chat	Enviar una pregunta y leer respuesta
Chat	Crear una nueva conversación
Chat	Borrar una conversación
Perfil	Ver carreras guardadas
Perfil	Expandir un tag del perfil
Perfil	Retomar quiz
Perfil	Borrar la cuenta

Nota. Elaboración propia.

Figura 45. Aspectos evaluados para la prueba del módulo de comunicación

Módulo	Pregunta	Aspecto UX/HCI Clave
Autenticación	¿Qué tan fluida, intuitiva y visualmente atractiva fue la interacción con el proceso de inicio de sesión y la gestión de tus datos personales?	Usabilidad, Estética, Fluidez
Quiz	¿La interfaz del quiz facilitó una interacción ágil y agradable, y la presentación de las preguntas contribuyó a una experiencia sin fricciones?	Interacción, Fluidez
Exploración	¿Qué tan estética, coherente y fácil de usar fue la interfaz para navegar y descubrir contenido, y la disposición visual de las recomendaciones te invitó a interactuar?	Estética, Coherencia, Interacción
Match	¿La representación visual de los niveles de afinidad fue clara y estéticamente agradable, y la interacción con las recomendaciones de carreras te resultó sencilla y directa?	Estética, Claridad, Interacción
Carreras	¿La presentación visual y la estructura de la información de las carreras y sus combinaciones fueron atractivas y fáciles de digerir, mejorando tu comprensión?	Estética, Organización, Usabilidad
Chat	¿La interfaz del chat fue estéticamente agradable y la interacción con él se sintió natural y eficiente para resolver tus dudas?	Estética, Interacción, Eficiencia
Perfil	¿La usabilidad de la sección de perfil permitió un acceso y una gestión fluida y clara de tus resultados y configuraciones de privacidad, con una interfaz coherente?	Usabilidad, Fluidez, Coherencia
General	Considerando todos los aspectos de la aplicación (estética, fluidez, interacción), ¿cuál es tu nivel de satisfacción general y qué tan probable es que la recomiendes?	Experiencia General, Estética, Fluidez

Nota. Elaboración propia.

Resultados de las pruebas

Figura 46. Métricas de la prueba 1 del módulo de comunicación

Primera Prueba	E2	E1	E0	#	Satisfacción	Comentarios
Autenticación	1	0	0	5		
Carreras	0	1	2	4		Animación para indicar que se puede ver más opciones de cursos. Arreglar error con pensums y año. El menú se encuentra un poco escondido
Chat	2	1	0	5		Usar otro ícono para revisar historial de chats, y en crear chat usar signo +
Exploración	1	0	0	4		
Match	2	0	0	4		
Perfil	4	0	0	5		
Quiz	2	0	0	5		

Nota. Elaboración propia.

Figura 47. Métricas de la prueba 2 del módulo de comunicación

Segunda Prueba 						
Módulo	E2	E1	E0	#	Satisfacción	Comentarios
Autenticación	1	0	0		4	
Quiz	2	0	0		5	
Exploración	1	0	0		4	Fue difícil entender las tarjetas de egresados, sería bueno agregar labels del tipo de tarjeta.
Match	2	0	0		5	Enviar el mensaje de una vez, en vez de colocarla como prompt
Carreras	3	0	0		4	
Chat	3	0	0		5	
Perfil	4	0	0		5	

Nota. Elaboración propia.

Figura 48. Métricas de la prueba 3 del módulo de comunicación

Tercera Prueba						
Módulo	E2	E1	E0	#	Satisfacción	Comentarios
Autenticación	1	0	0		4	
Quiz	2	0	0		5	
Exploración	1	0	0		4	Tener una vista del feed, en forma de cuadros
Match	2	0	0		5	Sería mejor poder resumir la información en imágenes, que en texto
Carreras	3	0	0		4	Elegir la carrera para combinar
Chat	3	0	0		5	Explicar más los resultados de sus tests en cuestión de la pregunta; No es tan claro el crear la conversación, por estar dentro de historial, se podría agregar el botón de crear a la par del de historial para hacerlo más claro; Ponerle nombre a los chats
Perfil	4	0	0		5	Efecto visual o tutorial para que las personas sepan que al presionar las tags les aparece más información; Opción de repetir examen dentro de la pantalla de carreras; El flujo de uso es muy ligero de información y no abruma con opciones y tutoriales

Nota. Elaboración propia.

Figura 49. Métricas de la prueba 4 del módulo de comunicación

Cuarta Prueba 									
Módulo	E0	E1	E2	#	Satisfacción	Comentarios			
Autenticación	1	0	0		3	No se ve bien el checkbox para aceptar condiciones			
Quiz	0	0	2		5	Empezar con las opciones positivas dentro de las preguntas. Las imágenes en algunas preguntas estan muy pequeñas			
Exploración	0	0	1		5				
Match	0	0	2		5				
Carreras	0	0	3		5				
Chat	0	0	3		4	No se sabe que se puede crear un nuevo chat a simple vista			
Perfil	0	0	4		5				

Nota. Elaboración propia.

Figura 50. Métricas de la prueba 5 del módulo de comunicación

Quinta Prueba 									
Módulo	E2	E1	E0	#	Satisfacción	Comentarios			
Autenticación	1	0	0		5				
Quiz	2	0	0		5	Que todas las opciones de likert estén en el mismo orden. Colocar indicador que se puede expandir la imagen para obtener más información. Indicar que existe un cronómetro para poder visualizar el enunciado			
Exploración	1	0	0		5				
Match	2	0	0		5				
Carreras	3	0	0		5				
Chat	3	0	0		4				
Perfil	4	0	0		5				

Nota. Elaboración propia.

Figura 51. Métricas de la prueba 6 del módulo de comunicación

Sexta Prueba						
Módulo	E2	E1	E0	#	Satisfacción	Comentarios
Autenticación	1	0	0		5	
Carreras	3	0	0		5	
Chat	3	0	0		5	
Exploración	1	0	0		4	
Match	2	0	0		4	
Perfil	4	0	0		5	
Quiz	2	0	0		5	Colocar las preguntas después de la explicación

Nota. Elaboración propia.