



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**PROPUESTA DE DESARROLLO PARA EL MODELO DE FORMACIÓN EN
INNOVACIÓN DE LA CARRERA DE INGENIERÍA CIVIL ELÉCTRICA,
USANDO HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL**

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL ELÉCTRICA

CARLA FRANCISCA GUZMÁN ROA

PROFESOR GUÍA:
ANDRÉS CABA RUTTE

MIEMBROS DE LA COMISIÓN:
FRANCISCO RIVERA SERRANO
SERGIO CELIS GUZMÁN

SANTIAGO DE CHILE
2024

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERA CIVIL ELÉCTRICA
POR: CARLA FRANCISCA GUZMÁN ROA
FECHA: 2024
PROF. GUÍA: ANDRÉS CABA RUTTE

PROPUESTA DE DESARROLLO PARA EL MODELO DE FORMACIÓN EN INNOVACIÓN DE LA CARRERA DE INGENIERÍA CIVIL ELÉCTRICA, USANDO HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL

La incorporación de información respecto al mercado laboral en la planificación de cursos y planes de formación universitaria es un tema relevante a discutir, puesto que la convergencia entre el mundo laboral y el mundo académico es necesaria. El mercado laboral está cambiando rápidamente, haciendo el análisis correspondiente más complejo.

Por otro lado, las herramientas de Inteligencia Artificial ya han sido usadas para fines educacionales. En este trabajo de título, se propone el uso de estas herramientas para el estudio del mercado laboral actual y del plan de formación en innovación de la carrera Ingeniería Civil Eléctrica de la Universidad de Chile. Con los hallazgos encontrados, se buscaron oportunidades de cambio para el plan curricular.

La metodología utilizada considera la recuperación de ofertas laborales dirigidas a estudiantes egresados de Ingeniería Civil Eléctrica en su variedad de especialidades. El filtrado de ofertas no relevantes para el estudio, en el que para decidir el mejor método se realizó un experimento de comparación de rendimiento en tareas de clasificación de dos modelos distintos. El trabajo continúa con la identificación de relaciones de los resultados de aprendizaje con las distintas ofertas de empleo. A partir de ese resultado, se realizó un análisis previo de resultados, terminando con la agrupación final de resultados de aprendizaje mediante tres métodos diferentes.

Los resultados obtenidos respaldan el uso de las herramientas de Inteligencia Artificial para el objetivo propuesto. Más aún, se genera una propuesta de cierre de brechas, considerando la aplicabilidad del resultado al contexto de la carrera estudiada.

Entre las principales conclusiones alcanzadas, se encuentra la validación del uso de herramientas de Inteligencia Artificial como GPT-4o o algoritmos de agrupación. Este trabajo de título demuestra que es posible aplicar IA a la construcción de planes de formación, en este caso, enfocados a la formación en innovación.

*A mí misma, por las noches que no dormí
y el esfuerzo de 6 años.*

Carla

Agradecimientos

En primer lugar me gustaría agradecer a mis papás, Mariela y Jorge, por nunca cortarme las alas y siempre creer que puedo lograr lo que me proponga. Tengo un poquito de los dos en mí, la fortaleza y determinación de mi mamá, y lo trabajador de mi papá, sin eso no lo habría logrado. Gracias por apoyarme en esta aventura de venir desde Chiguayante a Santiago, no fue fácil para ninguno, pero lo logramos.

A mi hermana Macarena, por siempre estar para mí, por ayudarme a ver las cosas más simples y siempre creer en mí. Por celebrar mis logros y acompañarme cuando fallé. Y por aguantar y entender mis mañas. Sin tu apoyo y los miles de empujoncitos que me diste, no habría llegado a este punto.

Quiero agradecer a mis abuelitos, Ulises y Marina, que aunque nunca entendieron lo que estudiaba, siempre estuvieron ahí para preguntar cómo iba, cómo estaba o si me hacía falta algo.

A mis amigos de la universidad. A Sofi y Vale por ser mis pilares en Santiago los primeros años, por siempre estar ahí en los momentos más duros de la carrera, por ir a cocinarme cuando yo no sabía, por los llantos y las risas (o ambos juntos), y por los buenísimos viernes. Y a Carlitos, nada que decir, no podría haber pedido mejor amigo.

A mis amigas del colegio. Nos perdimos una etapa larga de universidad, pero sin ustedes no habría llegado a donde estoy. Fueron las primeras personas fuera de mi familia en creer en mí y mis capacidades, me enseñaron a ser mejor persona y amiga, y pienso que sin ese valor que me transmitieron, no habría llegado acá.

A Antonia, mi apoyo fundamental en este proceso. Por estar conmigo siempre, por motivarme a seguir adelante, por tus comidas deliciosas y por siempre entender desde el amor. Este proceso fue mucho mejor gracias a ti.

Finalmente, muchas gracias al profesor Andrés Caba por darme la oportunidad de trabajar con él, por cuestionar y apoyar mis ideas, y siempre estar disponible durante todo este proceso. No podría haber trabajado con mejor profe guía.

A todos, gracias totales.

Carla Guzmán Roa

Tabla de Contenido

1. Introducción	1
1.1. Antecedentes generales	1
1.2. Identificación y formulación del problema	2
1.3. Objetivos	3
1.3.1. Objetivo general	3
1.3.2. Objetivos específicos	3
1.4. Organización de la memoria	3
2. Estado del Arte	4
2.1. Inteligencia artificial en planificación de cursos	4
2.2. Plan de formación en innovación actual	7
2.2.1. Cursos obligatorios	8
2.2.1.1. EL3105 - Seminario de Ingeniería Eléctrica e Innovación Tecnológica	8
2.2.1.2. EL4204 - Formulación y Evaluación de Proyectos	9
2.2.1.3. EL6101 - Taller de Proyectos Tecnológicos	9
2.2.1.4. EL6201 - Taller de Emprendimiento e Innovación	10
2.2.2. Cursos electivos	11
2.2.2.1. EL4030 - Seminario de Diseño e Innovación Tecnológica	11
2.2.2.2. EL5030 - Seminario de Diseño e Innovación Tecnológica II	12
2.2.2.3. EL6030 - Seminario de Diseño e Innovación Tecnológica III	12
2.3. Iniciativas activas en el DIE	12
2.3.1. Beauchef Proyecta	12
2.3.2. Impulsa DIE	14
2.3.2.1. Metodología Relámpago	14
3. Marco Teórico	16
3.1. <i>Web Scraping</i>	16
3.1.1. Código de una página web	16
3.1.2. Selenium	18
3.2. Natural Language Processing (NLP)	19
3.2.1. Large Language Models (LLM)	20
3.2.1.1. GPT-4o	23
3.2.1.2. BART	24
3.2.1.3. Clasificación <i>Zero-Shot</i>	24
3.2.2. Herramientas utilizadas para procesamiento de textos	25
3.2.2.1. Tokenización	25

3.2.2.2.	spaCy	25
3.2.2.3.	Análisis de Componente Principal (PCA)	26
3.3.	Clustering	26
3.3.1.	Agrupación basada en centroides	26
3.3.1.1.	K-Means	27
3.3.2.	Agrupación basada en la estructura jerárquica	28
4.	Etapas y Metodología del Proyecto	30
4.1.	Recolección de datos	31
4.1.1.	Análisis del currículum de Ingeniería Civil Eléctrica	31
4.1.2.	Recolección de ofertas de trabajo en la web	32
4.1.2.1.	Organización de la web trabajando.com	33
4.1.2.2.	<i>Web scraping</i>	34
4.2.	Preparación de los datos	35
4.2.1.	Unificación y limpieza de la base de datos	35
4.2.2.	Eliminación de registros no relevantes	36
4.2.2.1.	Identificación de registros no relevantes: Clasificación Zero-Shot	37
4.2.2.2.	Identificación de registros no relevantes: API GPT	37
4.2.2.3.	Evaluación de desempeño de los modelos	38
4.2.2.4.	Filtro final de la base de datos	38
4.3.	Etiquetado de la base de datos	39
4.3.1.	Clasificación con etiquetas determinadas	40
4.3.2.	Análisis preliminar de los resultados	42
4.4.	Construcción de la propuesta	42
4.4.1.	Preprocesamiento de los datos	43
4.4.2.	Aplicación de algoritmos de agrupación	44
4.4.2.1.	K-Means	44
4.4.2.2.	Agrupación Jerárquica	45
4.4.3.	Agrupación mediante API GPT-4o	45
4.5.	Consolidación de la propuesta	46
5.	Resultados	47
5.1.	Recolección de datos	47
5.1.1.	Análisis del currículum de Ingeniería Civil Eléctrica	47
5.1.2.	Recolección de ofertas de trabajo web	53
5.2.	Preparación de los datos	53
5.2.1.	Comparación de modelos de clasificación	53
5.2.2.	Eliminación de registros no relevantes	54
5.3.	Etiquetado de la base de datos	56
5.3.1.	Análisis preliminar de los resultados	56
5.4.	Construcción de la propuesta	59
6.	Discusión	64
6.1.	Análisis del currículum de Ingeniería Civil Eléctrica	64
6.2.	Preparación de los datos	64
6.2.1.	Comparación de modelos de clasificación	64
6.2.2.	Eliminación de registros no relevantes	65
6.3.	Etiquetado de la base de datos	66

6.3.1.	Análisis preliminar de los resultados	66
6.4.	Construcción de la propuesta	68
6.5.	Análisis final y consolidación de la propuesta	69
6.5.1.	K-Means	69
6.5.2.	Agrupación Jerárquica	71
6.5.3.	Agrupación con GPT-4o	73
6.5.4.	Consolidación de la Propuesta	76
7.	Conclusiones	78
7.1.	Trabajo Futuro	79
7.1.1.	Inclusión de otros cursos del plan de formación	79
7.1.2.	Identificar otras competencias a partir de herramientas de Inteligencia Artificial	80
7.1.3.	Inclusión de opinión de estudiantes	80
7.1.4.	Ajuste de pesos por SCT	80
	Bibliografía	81
	Anexos	84
	Anexo A. Resultados	84
A.1.	Descripción de Resultados de Aprendizaje	84
A.2.	Agrupación de Resultados de Aprendizaje	87
A.2.1.	Agrupación con GPT-4o (Respuesta)	87

Índice de Tablas

5.1.	Resultados análisis currículum universitario	47
5.2.	Organización en los distintos niveles: Grupo de competencia, Competencia, Subcompetencia y Resultado de Aprendizaje.	49
5.2.	Continuación. Organización en los distintos niveles: Grupo de competencia, Competencia, Subcompetencia y Resultado de Aprendizaje.	50
5.2.	Continuación. Organización en los distintos niveles: Grupo de competencia, Competencia, Subcompetencia y Resultado de Aprendizaje.	51
5.3.	Cantidad de ofertas de trabajo recuperadas por búsqueda realizada.	53
5.4.	Comparación de métricas de rendimiento de clasificación de textos con GPT-4o y <i>Zero-Shot</i>	54
5.5.	Comparación de métricas de rendimiento de clasificación de textos con GPT-4o y <i>Zero-Shot</i>	54
5.6.	Resultados del rendimiento del programa.	55
5.7.	Comparación cantidad de ofertas clasificadas por GPT 4o con el estado inicial de ofertas de empleo.	55
5.8.	Resultados del rendimiento del programa al identificar resultados de aprendizaje en ofertas de trabajo.	56
5.9.	Cantidad de ofertas de empleo en las que el modelo identifica cada resultado de aprendizaje.	57
5.10.	Análisis general del resultado del etiquetado.	58
5.11.	Nombre y enfoque de cursos entregados por GPT-4o.	61
5.12.	Clúster asignado a cada resultado de aprendizaje para los tres tipos de agrupación utilizada.	62
A.1.	Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.	84
A.1.	Continuación. Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.	85
A.1.	Continuación. Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.	86
A.1.	Continuación. Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.	87

Índice de Ilustraciones

2.1.	Mapa de competencias. Extraído de [9].	6
2.2.	Visualización de clusters. Extraído de [9].	6
2.3.	Esquemas implementados para alinear los cursos de Ingeniería Eléctrica, Ingeniería Mecánica y Diseño Industrial; los códigos corresponden a los cursos detallados en este capítulo. Extraído de [19, p. 5].	13
2.4.	Flujograma Metodología Relámpago. Extraído de [20].	15
3.1.	Tres distintas visualizaciones de una página web: La página, el código, el árbol DOM. Extraído de [23, p. 17].	17
3.2.	Ejemplo de obtención de información desde una página web basado en árbol DOM. Extraído de [23, p. 18].	17
3.3.	Ejemplo Chat GPT, conocimiento discursivo.	20
3.4.	Cálculo del valor a_3 , el tercer elemento de una secuencia utilizando el mecanismo de auto-atención. Extraído de [29, p. 219].	22
3.5.	Capa de autoatención casual (a) y capa de autoatención bidireccional (b). Extraído de [29, p. 243].	23
3.6.	Arquitectura de transformador usada en GPT. Extraído de [31, p. 4].	23
3.7.	Estrategias de transformadores. Derecha: estrategia usada en BERT. Izquierda: Estrategia usada en GPT. Extraído de [33, p. 2]	24
3.8.	Flujo de clasificación Zero-Shot (<i>bart-large-mnli</i>). Extraído de [34].	25
3.9.	Ubicación de centroides según iteraciones. Extraído de [39, 362].	28
3.10.	Estrategias de agrupación aglomerativa y divisiva. Extraído de [40, p. 460].	29
3.11.	Dendrograma típicamente usado para representar los resultados de agrupación jerárquica. Extraído de [40, p. 460].	29
4.1.	Etapas y Metodología trabajo de título.	30
4.2.	Organización de la oferta de empleo <i>trabajando.com</i>	33
4.3.	Árbol DOM elementos de interés en <i>trabajando.com</i>	34
5.1.	Relación Competencias Específicas - Curso - Resultado de Aprendizaje.	48
5.2.	Relación Competencias Genéricas - Curso - Resultado de Aprendizaje.	48
5.3.	Red de conexiones entre competencias, subcompetencias, resultados de aprendizaje y cursos bajo el grupo de competencias relacionadas al conocimiento.	51
5.4.	Red de conexiones entre competencias, subcompetencias, resultados de aprendizaje y cursos bajo el grupo de competencias relacionadas a las habilidades.	52
5.5.	Red de conexiones entre competencias, subcompetencias, resultados de aprendizaje y cursos bajo el grupo de competencias relacionadas a la actitud.	52
5.6.	Matrices de confusión resultantes de la clasificación con GPT 4o (izquierda) y <i>Zero-Shot</i> (derecha).	54
5.7.	Visualización del resultado generado por GPT 4o en la clasificación de ofertas relevantes para estudio, mostrado en la Tabla 5.3.	56

5.8.	Visualización de la cantidad de veces que una resultado de aprendizaje fue identificado en alguna oferta de trabajo.	58
5.9.	Visualización de la cantidad de ofertas que tienen la misma cantidad de resultados de aprendizaje identificados.	59
5.10.	Método del codo aplicado para determinar clústeres óptimos en K-Means.	60
5.11.	Resultados de agrupación con K-Means.	60
5.12.	Resultados de agrupación jerárquica aglomerativa.	61
5.13.	Visualización de los grupos creados por los diferentes métodos de agrupación. Izquierda: Clústeres creados por K-Means. Centro: Agrupación jerárquica. Derecha: Agrupación con GPT-4o.	63

Capítulo 1

Introducción

1.1. Antecedentes generales

Incorporar el análisis del mercado laboral a los planes de formación de carreras universitarias no es un desafío nuevo. Esta necesidad ha sido arduamente estudiada por importantes organizaciones como la OCDE, la cual destaca la importancia de ajustar continuamente los programas de formación educacional para reflejar las cambiantes demandas del mercado [1].

El desajuste entre las competencias incorporadas en los planes de formación universitaria y las necesidades reales del mercado laboral, pueden desembocar en mayores tasas de desempleo [2].

Recopilar y analizar datos precisos del mercado laboral es un desafío considerable, dificultando a las instituciones educativas ajustar los planes de formación para responder a las demandas del mercado laboral. Como se mencionó anteriormente, este problema se ha tratado de abordar desde distintas aristas, como en el artículo “Skills, Majors, and Jobs: Does Higher Education Respond?” de Johnathan G. Conzelmann et. al, donde se busca cuantificar el nivel de respuesta que tienen las instituciones de educación superior en Estados Unidos, a los cambios en las competencias buscadas en el mercado laboral [3].

La integración de las herramientas de Inteligencia Artificial en diferentes campos, han presentado esta alternativa como una buena opción para la mejora y optimización de procesos. La Inteligencia Artificial no se ha quedado fuera del área de la educación, incluyendo diversas aplicaciones como la personalización del aprendizaje y el análisis de datos educativos [4], [5].

La ingeniería, como disciplina, ha experimentado una transformación significativa en las últimas décadas. Ya no es suficiente dominar los conceptos teóricos y técnicos; se espera que los ingenieros sean innovadores, que puedan solucionar problemas de forma ágil y efectiva, y que sean capaces de colaborar con profesionales de distintas disciplinas para conseguir objetivos en común. En este contexto, la formación en innovación se convierte en un elemento crucial para preparar a los estudiantes no solo para los desafíos actuales, sino también para los que aún están por venir [6], [7].

El Departamento de Ingeniería Eléctrica (DIE) de la Universidad de Chile no ha ignorado esta necesidad, haciendo varios esfuerzos por mejorar este aspecto en la formación de los

estudiantes de la carrera de Ingeniería Civil Eléctrica, entre los cuales destaca un reciente cambio en la malla de estudios. El nuevo plan de formación cuenta con un marcado enfoque en el desarrollo de habilidades relacionadas a la innovación y, el perfil de egreso, considera estas nuevas habilidades y características en los profesionales egresados del DIE.

La iniciativa Impulsa DIE, es otra de las respuestas ante esta necesidad, el cual considera un apoyo a los emprendimientos y formación personalizada para estudiantes que se interesen en desarrollar proyectos propios.

A partir de esto, resulta a lo menos interesante preguntarse si se podrían usar herramientas de Inteligencia Artificial para proponer mejoras y cerrar brechas entre las competencias y habilidades esperadas respecto a innovación, en un egresado de esta carrera, y las que puedan ser aprendidas en el paso por la FCFM.

1.2. Identificación y formulación del problema

De acuerdo a los antecedentes revisados previamente, considerando especialmente la necesidad de realizar análisis de grandes cantidades de datos para generar información relevante para la construcción del plan de formación de la carreras ingeniería Civil Eléctrica, es que se propone la implementación de herramientas de Inteligencia Artificial para generar una propuesta de cierre de brechas entre el mercado laboral y las competencias actualmente incorporadas al plan de formación.

La rápida evolución de la tecnología y la creciente necesidad de competencias interdisciplinarias, plantean la pregunta central de cómo aportar al diseño de un plan de formación que sea ágil, pertinente y esté preparado para abordar las demandas cambiantes de la industria y la sociedad. Ante esto, resulta interesante utilizar herramientas de análisis de datos e Inteligencia Artificial para proponer posibles mejoras y ayudar a cerrar brechas entre, lo esperado por la industria y la formación otorgada por la universidad.

El enfoque del problema pretende ser una ayuda al análisis de la situación actual, considerando volúmenes de datos que serían difíciles de revisar de forma manual para los docentes del departamento encargados del diseño de los planes de formación. Esta es la principal motivación para el desarrollo de este trabajo.

Es importante mencionar que el uso de herramientas de Inteligencia Artificial no apunta a reemplazar el trabajo hecho por los docentes, sino que a generar una herramienta que, en conjunto a la expertiz de los profesores, sea capaz de mejorar el plan de formación en innovación de la carrera Ingeniería Civil Eléctrica.

1.3. Objetivos

1.3.1. Objetivo general

El objetivo general de esta memoria es generar una propuesta de desarrollo para el modelo de formación en Innovación de la carrera de Ingeniería Civil Eléctrica a partir de análisis de datos cuantitativos y cualitativos usando herramientas de inteligencia artificial.

1.3.2. Objetivos específicos

Para poder llevar a cabo el objetivo general de esta investigación, se proponen los siguientes objetivos específicos:

- Recuperar información cualitativa y cuantitativa respecto a las habilidades y competencias entregadas en la carrera, así como las buscadas en el mercado.
- Preprocesar los datos recolectados y aplicar algoritmos de *clustering* para agrupar habilidades y competencias.
- Analizar el resultado del *clustering* en base al modelo de formación actual y la experiencia de profesores dedicados al tema para desarrollar la propuesta final.

1.4. Organización de la memoria

Esta memoria de título se organiza en capítulos. A continuación, se listan junto a una breve descripción de cada uno.

1. Capítulo 1. Introducción: en este capítulo se presentan los antecedentes, la identificación del problema y los objetivos de este trabajo de título.
2. Capítulo 2. Estado del Arte: se presentan otras investigaciones que utilizan herramientas de Inteligencia Artificial para el análisis de planes de formación. Además, se hace revisión de las iniciativas actuales del plan de formación en innovación en Ingeniería Eléctrica.
3. Capítulo 3. Marco Teórico: se exponen los principales aspectos teóricos necesarios para entender y respaldar la metodología utilizada.
4. Capítulo 4. Etapas y Metodología del Proyecto: en este capítulo se explican todos los pasos llevados a cabo en esta investigación.
5. Capítulo 5. Resultados: se presentan los resultados obtenidos en el mismo orden propuesto en la metodología de esta memoria de título.
6. Capítulo 6. Discusión: en este capítulo se describen y analizan los resultados presentados en el capítulo anterior, siguiendo el mismo orden propuesto inicialmente.
7. Capítulo 7. Conclusiones: finalmente, se exponen las conclusiones obtenidas y se incluye la propuesta de trabajo futuro.

Capítulo 2

Estado del Arte

En esta sección se hace una revisión del estado del arte del presente trabajo de título, separada en 3 áreas principales. En primer lugar, se hace una revisión de aplicaciones en las que ya se haya utilizado la Inteligencia Artificial para la planificación de cursos. Se revisa el plan de formación en innovación actual de la carrera Ingeniería Civil Eléctrica del Departamento de Ingeniería Eléctrica de la Universidad de Chile y, por último, se ahonda en las iniciativas enfocadas en innovación y emprendimiento activas en el DIE.

2.1. Inteligencia artificial en planificación de cursos

No son abundantes los estudios referidos al uso de herramientas de Inteligencia Artificial como apoyo en la planificación de cursos, pero no es inédito.

En Septiembre del 2023 Hyeji Yang et. al, publican un estudio titulado “Analyzing the Alignment between AI Curriculum and AI Textbooks through Text Mining”[8]. En este documento, se realiza una revisión a la alineación entre los textos utilizados para la formación en IA en los ciclos escolares en Corea del Sur y el currículum diseñado para la enseñanza de la Inteligencia Artificial, con el objetivo de proponer mejoras en el plan de formación.

Se detalla la metodología de la investigación:

1. Revisión de investigaciones relacionadas al tema, incluyendo extensa bibliografía. En particular se puede resumir este proceso como:
 - Revisión del distintos programas de educación en Inteligencia Artificial, tanto en Corea del Sur como en otros países (India y China).
 - Comprensión de las formas de analizar textos educativos.
 - Revisión de bibliografía referida al análisis de texto utilizando técnicas de Minería de Textos.
2. Definición de la metodología de investigación.
 - Se explica cómo se llevó a cabo el proceso de la recolección de datos para el trabajo.
 - Revisión del preprocesamiento de los datos para el análisis.

- Explicación del método de análisis.
3. Resultados. En esta etapa se presentan los resultados por etapa tras aplicar el algoritmo, con el objetivo de estructurar la data para proponer mejoras finales.
 4. Presentación de Discusión y Conclusiones.

De los puntos mencionados anteriormente, es relevante para esta memoria destacar la revisión bibliográfica realizada respecto al análisis de textos utilizando técnicas de Minería de Textos. En este trabajo se propone hacer una mezcla entre los métodos *Term Frequency-Inverse Document Frequency (TF-IDF)* y *Latent Dirichlet Allocation (LDA)*. Hyeji Yang et. al, mencionan que ambas técnicas se han utilizado para hacer análisis de textos educativos, pero hasta el momento no se habían combinado para un único análisis [8].

Otro aspecto relevante, es la forma de organizar la información para ser analizada. En primer lugar, se identificaron en el currículum vigente 4 áreas de formación específicas. A partir de esto, se organizó la información en una tabla donde se identificaron conceptos claves, contenidos y resultados de aprendizaje para cada una de las áreas.

Como segundo paso del proceso de recolección de datos, se llevó a cabo una revisión para determinar la importancia de cada una de las áreas en los 8 textos que fueron analizados. Esto hizo mediante el número de páginas dedicadas a cada área.

El preprocesamiento de datos se llevó a cabo utilizando KoNLPy, una biblioteca de análisis del lenguaje Coreano. Con esto, se logró la “tokenización” de los datos.

Para realizar el estudio se hizo un análisis de frecuencias (TF) para cada texto, y para cada una de las áreas, seleccionando los 20 términos más frecuentes y relacionarlos con los resultados de aprendizaje identificados previamente. A continuación se utilizó TF-IDF para determinar las 15 palabras con valores más altos, con el objetivo de poder relacionar estas palabras con la informática.

Para la visualización de los datos se escogió la librería PyLDAvis. Librería especializada en la comprensión de análisis de minería de textos. Con esto se visualizó la composición temática de cada libro, para así compararlo con el currículum vigente.

El estudio concluye de manera exitosa, entregando directrices para las mejoras al currículum de formación en Inteligencia Artificial; y también mejoras a los textos.

Este trabajo referenciado, resulta útil para el objetivo de esta memoria, puesto que se enfoca en hacer un cierre de brechas entre los textos utilizados para formar estudiantes en un tema determinado (Inteligencia Artificial), y el currículum vigente en Corea del Sur relacionado.

Por otro lado, en Agosto del 2014 Christos Vaitzis, Christos et. al, publican el documento “Big Data in Medical Informatics: Improving Education Through Visual Analytics”[9]. En este estudio lo que se revisa son distintas formas de visualizar las competencias respecto al plan de formación de estudiantes de medicina. En particular, se revisan 2 tipos de visualizaciones:

- Mapa de competencias

En la Figura 2.1 se puede observar un mapa de competencias hecho a partir de los datos de un plan de formación de estudiantes de medicina. En verde oscuro se presentan 3 competencias principales o claves identificadas. A partir de estas tres competencias, se definen teóricamente 16 sub-competencias asociadas, que se muestran en color verde claro. En la parte inferior de la imagen, los puntos rojo oscuro representan las 8 competencias principales de un nivel en específico de formación estudiado y, en rojo claro al centro, se presentan distintos resultados de aprendizajes pertenecientes al programa. Esta visualización resulta bastante útil para apreciar la relación que hay entre competencias del plan de formación y los resultados de aprendizaje.

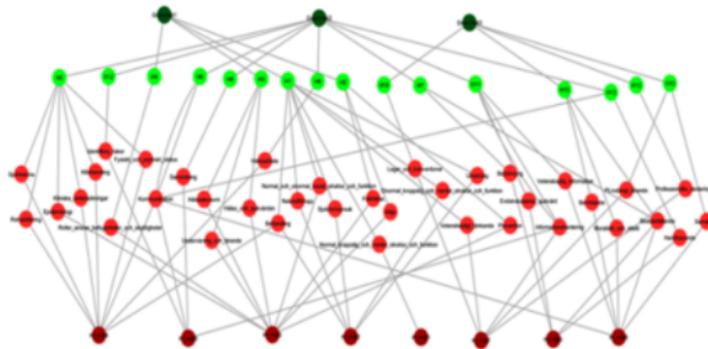


Figura 2.1: Mapa de competencias. Extraído de [9].

- Agrupación de competencias

La representación de colores para esta visualización es la misma que para el mapa de competencias. En este caso utilizaron métodos de agrupación, obteniendo 4 grupos principales. 3 de ellos están centrados en las 3 competencias principales, respectivamente y, el cuarto (de gris), agrupa las competencias referentes al nivel específico de formación.

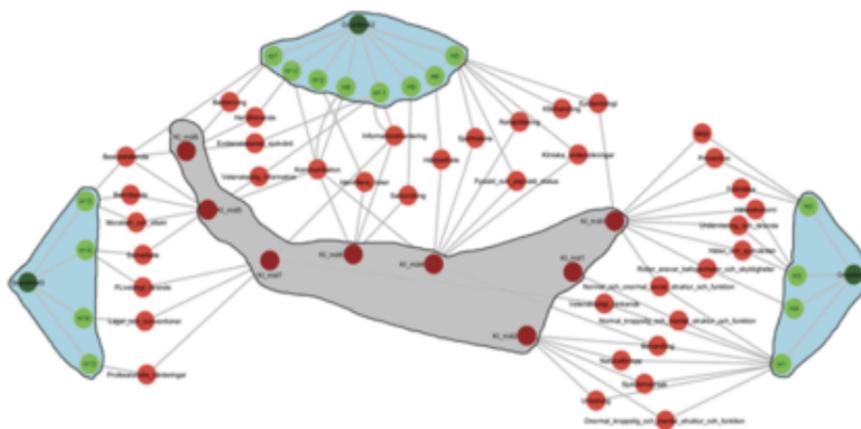


Figura 2.2: Visualización de clusters. Extraído de [9].

Las dos formas de visualizar la información extraídas del estudio son útiles y pertinentes a este trabajo, ya que apuntan a la construcción de un plan de formación relacionado con resultados de aprendizaje esperados.

2.2. Plan de formación en innovación actual

El año 2019 se concretó un importante cambio en el plan de estudio de las carreras de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Esta actualización apuntaba a reducir la duración del plan de estudio y actualizar los contenidos, para así acercar a las carreras impartidas por la FCFM a los estándares internacionales. En palabras de McPhee, entonces vicedecano de la facultad y director del Proyecto Ingeniería y Ciencias 2023, “La modificación de la nueva malla curricular de las licenciaturas y carreras que ofrece la Facultad, nos dio la oportunidad de revisar y ajustar nuestros perfiles de egreso, agregando conceptos importantes como la sustentabilidad y la innovación” [10, p. 82].

Es así, como los nuevos planes de estudio en la FCFM, apuntan a obtener un perfil de ingeniero especializado y técnico, pero que a la vez considere nuevas capacidades como la invención, innovación y emprendimiento. Es importante también para el perfil de egreso incluir las capacidades de comunicación efectiva en todas sus aristas (oral, escrita y gráfica) [10].

La carrera impartida por el Departamento de Ingeniería Eléctrica no quedó exenta de estas modificaciones. El nuevo perfil de egreso para la carrera reafirma el enfoque hacia el desarrollo de las habilidades de innovación y emprendimiento tecnológico mencionado anteriormente. En particular, se menciona de manera textual en la documentación disponibilizada por la Vicerrectoría de Asuntos Académicos de la FCFM, “(...) Asimismo, es capaz de concebir, implementar y gestionar proyectos tecnológicos considerando requerimientos económicos, ambientales, sociales y éticos, logrando soluciones eficientes, viables y con alto valor agregado.”[11].

Además, dentro del mismo perfil, al describir habilidades específicas se menciona “(...) Concebir, implementar y gestionar proyectos tecnológicos en áreas de la Ingeniería Eléctrica tales como automatización y supervisión de procesos industriales, sistemas de generación, transmisión y distribución de energía eléctrica, tecnologías de la información y la comunicación, procesamiento de señales, instrumentación, inteligencia computacional y robótica; identificando necesidades presentes y futuras y considerando para estos fines aspectos técnicos, económicos, sociales, ambientales, y aspectos legales y regulatorios”[11].

Al considerar las competencias transversales del perfil de egreso del ingeniero eléctrico de la Universidad de Chile, se encuentra específicamente la innovación como “Concebir ideas viables y novedosas para resolver problemas o necesidades, materializadas en productos, servicios o en mejoras a procesos, considerando el contexto sociocultural, económico y los beneficios para el usuario”[11].

Finalmente, dentro de los requisitos que debe cumplir el egresado de la carrera, se detalla la capacidad de “desarrollar tanto la capacidad de invención, innovación y emprendimiento, como el pensamiento crítico”[11].

Al hablar específicamente de formación en innovación en la carrera de Ingeniería Eléctrica, el principal cambio que se implementó fue la unión de los cursos obligatorios de especialidad EL5001 Introducción al Taller de Proyecto, EL5002 Introducción al Taller de Diseño, EL5003 Taller de Proyecto y EL5004 Taller de Diseño, en un sólo nuevo gran curso, EL6101 Taller de Proyectos Tecnológicos. Es decir, se resumen 4 cursos de 6 SCT, en uno de 12 SCT.

Con esto, la oferta académica actual del Departamento de Ingeniería Eléctrica dirigida al desarrollo de habilidades de innovación, emprendimiento y desarrollo de proyectos es:

Cursos obligatorios:

- EL3105 - Seminario de Ingeniería Eléctrica e Innovación Tecnológica
- EL4204 - Formulación y Evaluación de Proyectos
- EL6101 - Taller de Proyectos Tecnológicos
- EL6201 - Taller de Emprendimiento e Innovación

Cursos electivos:

- EL4030 - Seminario de Diseño e Innovación Tecnológica
- EL5030 - Seminario de Diseño e Innovación Tecnológica II
- EL6030 - Seminario de Diseño e Innovación Tecnológica III

Estos cursos tributan a ciertas competencias específicas (CE) y genéricas (CG), relacionadas al perfil de egreso de la carrera de Ingeniería Civil Eléctrica. Estas, acompañadas de la descripción de cada una, se añaden en el Anexo A de este documento.

A continuación, se ofrece una breve descripción de los cursos antes mencionados, en base a lo que propone el programa de cada uno de ellos.

2.2.1. Cursos obligatorios

2.2.1.1. EL3105 - Seminario de Ingeniería Eléctrica e Innovación Tecnológica

Objetivo general del curso: este curso tiene como propósito que los estudiantes de Ingeniería Eléctrica conozcan las distintas áreas de desarrollo en la carrera. Además, busca que los y las estudiantes analicen un desafío propuesto, diseñen e implementan un prototipo de baja resolución, como propuesta de valor de carácter conceptual, que demuestre total o parcialmente la solución a un problema, considerando el uso de software, hardware, modelos o cualquier soporte audiovisual [12].

Resultados de aprendizaje: dentro de los resultados de aprendizaje propuestos en el programa de este curso se destacan los que tienen mayor relación con la innovación.

- RA2: analiza un desafío propuesto y el contexto en que se inserta (cultural, social ambiental y económico), considerando las etapas del proceso de innovación para proponer ideas y seleccionar aquella que pueda transformarse en un potencial proyecto plasmable en un prototipo de baja resolución [12].

- RA3: define el problema, a partir del desafío propuesto, con su correspondiente justificación, gestionando la información del usuario, entorno y ecosistema, a través de un análisis de lo existente de la industria, a fin de identificar potenciales oportunidades del proyecto y su impacto [12].
- RA4: diseña e implementa, con criterio de innovación, un prototipo de baja resolución que demuestre total o parcialmente la solución de carácter tecnológico para un problema y su potencial impacto, plasmable en diferentes formatos como software, hardware, modelos o cualquier soporte audiovisual [12].
- RA7: evalúa la propuesta de solución al desafío planteado, considerando metodologías de innovación, así como un análisis de contexto en que este se inserta, y aspectos técnicos, económicos y de factibilidad, entre otros [12].

Adicionando además, resultados de aprendizaje referentes a la producción de documentos escritos y a la presentación de resultados con apoyo de recursos audiovisuales.

Se puede concluir que el curso está enfocado a la primera fase de diseño de un proyecto, que es identificar y definir un problema, para así proponer una solución.

2.2.1.2. EL4204 - Formulación y Evaluación de Proyectos

Objetivo general del curso: el curso tiene como propósito que los y las estudiantes formulen y evalúen proyectos de inversión públicos y privados en ingeniería eléctrica, con sus respectivas etapas y componentes, integrando a su formulación criterios técnicos, económicos, sociales, ambientales y éticos [13].

Resultados de aprendizaje: dentro de los resultados de aprendizaje propuestos en el programa de este curso, se destacan los que tienen mayor relación con la innovación.

- RA1: formula proyectos de inversión públicos y privados en ingeniería eléctrica, considerando etapas y componentes e integrando a su formulación criterios técnicos, económicos, sociales, ambientales y éticos [13].
- RA2: calcula indicadores de rentabilidad de un proyecto en ingeniería eléctrica, desde una perspectiva privada y social, a fin de determinar su viabilidad [13].
- RA3: evalúa proyectos de inversión públicos y privados en ingeniería eléctrica a partir del análisis de criterios e indicadores respecto de sus beneficios y costos en materias técnicas, económicas, ambientales, sociales y éticas [13].

Se puede concluir que este curso está completamente enfocado a la evaluación de proyectos de innovación con mirada al cálculo de indicadores de rentabilidad.

2.2.1.3. EL6101 - Taller de Proyectos Tecnológicos

Objetivo general del curso: el propósito del curso es desarrollar un proyecto tecnológico dentro de alguna de las cinco áreas de especialización de la carrera de Ingeniería Eléctrica (Energía, TICS, Control de Sistemas, Instrumentación e Inteligencia computacional y Robótica). El objetivo final del proyecto será definido ya sea por el cuerpo académico en conformidad del equipo de estudiantes, o por el equipo de estudiantes y ratificado por el cuerpo docente

[14].

Resultados de aprendizaje: dentro de los resultados de aprendizaje propuestos en el programa de este curso, se destacan los que tienen mayor relación con la innovación.

- RA2: diseña y proyecta una solución a un problema, utilizando metodologías ágiles para productos, ingeniería de proyecto de innovación y de integración, y para proyectos tradicionales, a fin de agregar valor a un sistema o proceso que permita mejoras al bienestar humano [14].
- RA3: evalúa tiempos, costos de procesos productivos de piezas y componentes eléctricos, considerando tecnologías y procesos de manufactura disponibles, diseño de prototipos, experimentos e información de mercado, etc., para proponer opciones de mejora [14].
- RA4: evalúa la factibilidad económica de cada etapa del estudio de ingeniería y/o el desarrollo de un producto, considerando las condiciones de mercado y licitaciones posibles y sugiriendo mejoras de ser necesario [14].
- RA5: elabora documentación técnica, planos y memorias de cálculo, considerando rigurosidad en el manejo de estándares y buenas prácticas asociadas a la ética profesional en ingeniería, los que integra a la propuesta de proyecto en cada una de las áreas de especialidad [14].

En general, se puede concluir que este curso está más enfocado al proceso de desarrollo de una innovación tecnológica, considerando evaluar tiempos de duración de las tareas, costos de desarrollo, factibilidad económica, generación de documentación técnica de un producto. Estos pasos indican que el curso se enfoca en la organización del equipo de trabajo con el objetivo de generar un Producto Mínimo Viable (MVP) al final del ramo.

2.2.1.4. EL6201 - Taller de Emprendimiento e Innovación

Objetivo general del curso: en el programa del curso se declara el objetivo general del curso con las siguientes palabras: “El curso tiene como propósito que el estudiantado desarrolle un Plan de Negocios de un proyecto, innovador y original, considerando necesidad o dolor comercial o social, oportunidad y tamaño del problema, así como una evaluación económica de este proyecto, iterando este plan con técnicas como *Design Thinking*, elaboración de maquetas y uso de Canvas”[15].

Resultados de aprendizaje: este curso se divide en 4 fases, las cuales tienen asociadas distintos resultados de aprendizaje. Entre ellos se destacan algunos a continuación:

- Fase 1: Planificación. Identifica un dolor o necesidad de la sociedad y de las personas sobre el cual se trabajará, a través de una prospección del problema, considerando mercado actual y sus variaciones futuras, opciones de financiamiento, impacto social, cultural y económico [15].
- Fase 1: Planificación. Define herramientas de emprendimiento con las cuales trabajará el dolor o necesidad, por ejemplo, *Design Thinking*, maquetas (*mock up*) y canvas, entre otros [15].
- Fase 2: Aprendizaje y descubrimiento. Levanta evidencia que sustente la estimación, el mercado en torno al cliente y tamaño de este, considerando, por ejemplo, espacio

geográfico, tendencias, barreras de entrada, entre otros, para ello, revisar las definiciones y técnicas tales como TAM (mercado total), SAM (mercado objetivo) y SOM (mercado obtenible) [15].

- Fase 2: Aprendizaje y descubrimiento. Construye la primera versión del plan de negocios, mediante herramientas como, por ejemplo, *Design Thinking*, maquetas (*mock up*), Lean Launchpad, las metodologías de Business Model Canvas o Lean Canvas, Modelo de desarrollo de clientes y metodologías ágiles, considerando los recursos clave y su disponibilidad [15].
- Fase 3: Generación de alternativas. Testea o prototipa el mínimo producto viable y su demanda, considerando prueba y error/*design thinking*, mercado, definición y dinámica del producto o servicio e identificando puntos de mejora y posibles inconvenientes que puedan surgir [15].
- Fase 4: Elabora un Plan de Negocios, para la escalabilidad y replicabilidad del negocio, considerando los siguientes aspectos: 1. un plan comercial, 2. un plan técnico, 3. un plan financiero y 4. un plan de desarrollo de productos [15].

En general, se puede concluir que este curso tiene un enfoque en el proceso de diseño de un proyecto innovador. Se identifica un dolor, se itera sobre la validación con el usuario, uso del modelo Canvas y se generan alternativas para suplir las posibles deficiencias del proyecto observadas mediante las validaciones y conversaciones con usuarios que se ven afectados directamente por el dolor al que apunta el proyecto. Todo lo anterior con miras a desarrollar un plan de negocios para un posible emprendimiento.

Es importante mencionar que en este curso se trabaja con maquetas o *mock up*, no se desarrolla el proyecto, como en el curso EL6101.

2.2.2. Cursos electivos

El Departamento de Ingeniería Civil Eléctrica ofrece una línea de cursos consecutivos relacionados al desarrollo de proyectos. Las opciones que componen esta línea son EL4030, EL5030 y EL6030. Son de carácter electivo en la malla de Ingeniería Eléctrica, y son cursos consecutivos (requisito del curso siguiente).

Los tres cursos ofrecidos por el DIE en esta línea basan su desarrollo en la metodología CDIO (Concebir, Diseñar, Implementar y Operar), enfocándose cada uno en distintas etapas de ella.

2.2.2.1. EL4030 - Seminario de Diseño e Innovación Tecnológica

Objetivo general del curso: este curso está diseñado para ser una introducción al enfoque CDIO. El objetivo es que los estudiantes logren comprender y conocer el enfoque CDIO.

Resultado de aprendizaje: se proponen 3 resultados de aprendizaje en el programa del curso.

- Los estudiantes comprendan el enfoque CDIO [16].
- Los estudiantes apliquen en un proyecto real que requiera de un alto nivel técnico y de innovación los principios del enfoque CDIO [16].

- Los estudiantes adquieran habilidades para trabajar en equipos multidisciplinarios [16].

Se puede concluir de este programa, que el curso quiere dar a conocer el enfoque CDIO, con el objetivo de que al continuar con la línea de cursos se pueda tener una visión general de todas las etapas antes de comenzar con las especificaciones y trabajo en cada una.

2.2.2.2. EL5030 - Seminario de Diseño e Innovación Tecnológica II

Objetivo general del curso: el objetivo general de este curso es implementar en detalle las dos primeras etapas del enfoque CDIO, que corresponden a C: Concebir, y D: Diseñar.

Resultado de aprendizaje: se proponen 3 resultados de aprendizaje en el programa del curso.

- Analiza situaciones profesionales, identificando sus dilemas y conflictos éticos, valorando las consecuencias de las decisiones en las distintas áreas de la ingeniería Eléctrica [17].
- Analiza soluciones a problemas específicos para un proyecto ingenieril dentro de un contexto multidisciplinario[17].
- Analiza procesos creativos para planificar estrategias de innovación para una resolución de un problema [17].

A modo de resumen del curso, se puede decir que está pensado para hacer un breve recuerdo del enfoque CDIO, pero la parte más relevante del curso apunta a la concepción y diseño de un proyecto innovador y tecnológico.

2.2.2.3. EL6030 - Seminario de Diseño e Innovación Tecnológica III

Objetivo general del curso: el objetivo general de este curso es implementar en detalle las dos últimas etapas del enfoque CDIO, que corresponden a I: Implementar, y O: Operar.

Resultado de aprendizaje: se proponen 3 resultados de aprendizaje en el programa del curso:

- Evaluar concepciones y diseños de proyectos [18].
- Aplicar los principios de implementación y operación en un proyecto real de largo plazo que requiera de un alto nivel técnico y de innovación [18].
- Generar habilidades, conocimientos y herramientas para implementar y operar proyectos en equipos numerosos y multidisciplinarios [18].

Se puede concluir que este curso está enfocado en la implementación y operación de un proyecto tecnológico. Dado este enfoque, es imperante que los proyectos tengan una base sólida, es decir, haber realizado previamente los pasos de concepción y diseño del producto.

2.3. Iniciativas activas en el DIE

2.3.1. Beauchef Proyecta

Beauchef Proyecta, según Matamala, “se concibe como una unidad de la FCFM que apunta al contribuir al desarrollo de proyectos multidisciplinarios con un foco en innovación tecnológica, que sean significativos a la formación de ingenieras e ingenieros líderes para la sociedad”

[19, p. 3]. Esta unidad se ha encargado desde el año 2016 de implementar la formación en innovación desde una mirada multidisciplinaria, tratando de crear lazos entre distintas carreras y facultades para el desarrollo de proyectos.

La iniciativa comienza como un desarrollo multidisciplinario entre los departamentos de Ingeniería Eléctrica e Ingeniería Mecánica, llegando a involucrarse en distintos cursos perteneciente a las distintas etapas de la formación ingenieril de la facultad (plan común, electivos de licenciatura, ramos obligatorios de especialidad y memorias multidisciplinarias).

Tras casi tres años de iteraciones y propuestas, hacia el final del 2018 Beauchef Proyecta contaba con una sólida propuesta de estructuración para los cursos, específicamente para las especialidades de Ingeniería Civil Mecánica e Ingeniería Civil Eléctrica, junto a estudiantes de la carrera de Diseño Industrial de la FAU y a estudiantes de periodismo de la universidad.



Figura 2.3: Esquemas implementados para alinear los cursos de Ingeniería Eléctrica, Ingeniería Mecánica y Diseño Industrial; los códigos corresponden a los cursos detallados en este capítulo. Extraído de [19, p. 5].

La Figura 2.3 muestra las iteraciones que han tenido la estructuración y alineación de los distintos cursos propuesto por Beauchef Proyecta.

Del trabajo realizado durante esos años por Beauchef Proyecta, se obtienen destacables conclusiones a partir de los comentarios hechos por los mismos alumnos en las encuestas docentes y conversaciones. Se pudo detectar que los cursos de la línea de especialidad eran los que más problemas percibían los estudiantes, principalmente por la falta de tiempo para desarrollar proyectos y el poco conocimiento de técnicas para desarrollarlos correctamente (organización y planificación) [19].

En el último tiempo los cursos con participación de Beauchef Proyecta se han visto sometidos a cambios importantes, principalmente por el reciente cambio de malla y duración de las ingenierías en la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, lo que no significa que la experiencia recolectada desde el 2016 haya sido perdida.

Poniendo el foco en los cursos electivos y obligatorios pertenecientes al Departamento de Ingeniería Eléctrica, Beauchef Proyecta, junto a Impulsa DIE, participa activamente del desarrollo de estos. En el contexto del cambio de malla, la formación obligatoria en la carrera combinó los cursos EL5001 Introducción al Taller de Proyecto, EL5002 Introducción al Taller de Diseño, EL5003 Taller de Proyecto y EL5004 Taller de Diseño, en un sólo nuevo gran curso, EL6101 Taller de Proyectos Tecnológicos, descrito anteriormente. Es mencionado en esta sección, puesto que los cursos EL5002 y EL 5004 eran cursos centrales en el plan propuesto

por Beauchef Proyecta.

2.3.2. Impulsa DIE

Impulsa DIE es una iniciativa estudiantil pensada para potenciar la innovación y el emprendimiento dentro del Departamento de Ingeniería Eléctrica. Actualmente, cuentan con participación dentro de 2 cursos obligatorios y 3 cursos electivos del departamento. Estos son:

Cursos obligatorios:

- EL6101 - Taller de Proyectos Tecnológicos
- EL6201 - Taller de Emprendimiento e Innovación

Cursos electivos:

- EL4030 - Seminario de Diseño e Innovación Tecnológica
- EL5030 - Seminario de Diseño e Innovación Tecnológica II
- EL6030 - Seminario de Diseño e Innovación Tecnológica III

Dentro de estos cursos el rol de Impulsa DIE es facilitar y proponer actividades que potencien la innovación y el emprendimiento. En particular, Impulsa DIE se ha encargado de concretar relaciones con empresas reales y activas en la industria, con el objetivo de articular el trabajo universidad-empresa.

2.3.2.1. Metodología Relámpago

Dentro del trabajo que realiza Impulsa DIE, se destaca la confección de la llamada Metodología Relámpago. Esta metodología trata de implementar el aprendizaje constante en el proceso de innovación y emprendimiento.

En la Figura 2.4 se presenta la visión general de la Metodología Relámpago propuesta por la iniciativa Impulsa DIE.

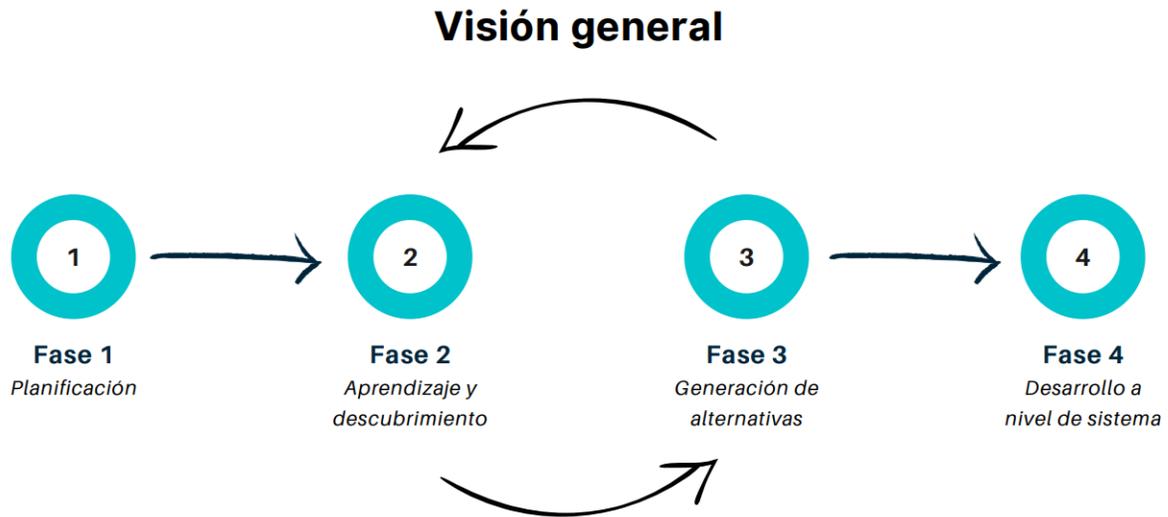


Figura 2.4: Flujograma Metodología Relámpago. Extraído de [20].

Tal como se puede apreciar en la Figura 2.4 la Metodología Relámpago consta de 4 fases:

1. **Planificación:** esta fase considera la definición del dolor sobre el cual se trabajará, acompañada de la planificación del trabajo a desarrollar. Dentro de un equipo esto contempla elección de la metodología de trabajo, la definición de roles y la definición de tareas claves para el desarrollo incluyendo la estimación de los tiempos que tomará desarrollar cada una de ellas.
2. **Aprendizaje y descubrimiento:** se basa en la comprensión de la problemática a resolver. En esta etapa el equipo se centra en identificar fortalezas y debilidades del proyecto, y necesidades latentes, para poder ser clasificadas según la importancia. Se plantea la investigación a fondo de la temática en estudio con tal de lograr redefinir el problema inicial en caso de ser necesario. También es posible crear sub-equipos de trabajo si es que se considera óptimo.
3. **Generación de alternativas:** en esta fase se generan ideas usando metodologías adecuadas (lluvia de ideas, mapas mentales, etc), se evalúan y categorizan, seleccionando una como la más atractiva. El objetivo es testear si la idea seleccionada como más atractiva resuelve la problemática, o si es necesario volver al paso anterior para redefinir el proyecto e iniciar con una nueva idea.
4. **Desarrollo a nivel de sistema:** esta es la etapa donde finalmente se implementa la iniciativa. El foco está puesto en ejecutar la idea, aprender de ella y refinar los detalles que sean necesarios, mediante el método SCAMPER (Sustituir, Combinar, Adaptar, Modificar, Propósito, Eliminar (o minimizar) y Reinvertir).

Es importante mencionar que esta metodología propuesta por Impulsa DIE está inspirada en el documento *El Arte de Fallar*, de Carlos A. Osorio [21].

Capítulo 3

Marco Teórico

3.1. *Web Scraping*

Para poder recopilar data desde sitios de internet hay dos métodos posibles: hacerlo de forma manual, lo cual resulta una tarea tediosa y poco eficiente, o utilizar herramientas que permitan automatizar la recolección de datos desde páginas web para así transformarla en una base de datos estructurada. El *web scraping* se refiere precisamente a eso: es una forma de automatizar la recolección de datos alojados en internet para construir bases de datos estructuradas [22].

El *web scraping* se puede realizar utilizando una gran variedad de herramientas disponibles, dependiendo de cual sea el lenguaje de programación escogido. Para este trabajo de título se ahondará en las librerías disponibles para realizar esta técnica en Python.

3.1.1. Código de una página web

Los documentos web están organizados siguiendo una estructura de árbol, llamada Modelo de Objetos del Documento, o en inglés *Document Object Model (DOM)* [23].

Como se muestra en la Figura 3.1, se pueden observar las páginas web desde tres distintos puntos de vista: documento mostrado en la web, código (en el ejemplo es HTML) y finalmente el DOM asociado a la página.

Estos documentos son recorridos por los programas de *web scraping* siguiendo la ruta que propone el usuario, ayudándose de selectores CSS o XPATHS. En la Figura 3.2 se puede observar un ejemplo de cómo se utilizan los árboles DOM para organizar la búsqueda de información y acceder a ella. En el ejemplo, se puede ver en la línea color rojo el camino determinado (XPATHS) para acceder a la información en el objeto final.

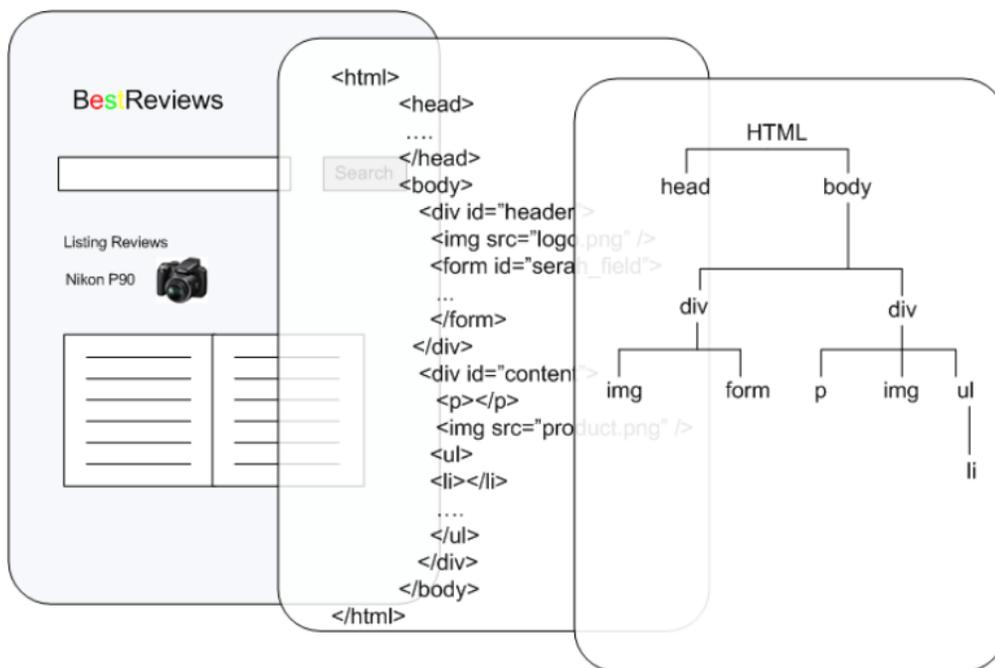


Figura 3.1: Tres distintas visualizaciones de una página web: La página, el código, el árbol DOM. Extraído de [23, p. 17].

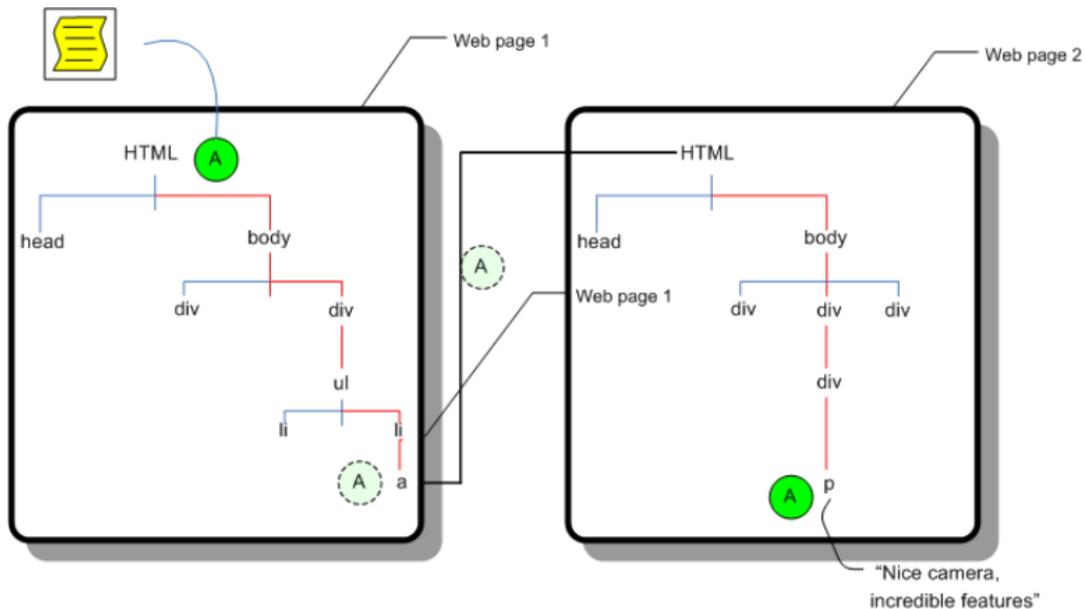


Figura 3.2: Ejemplo de obtención de información desde una página web basado en árbol DOM. Extraído de [23, p. 18].

Para recuperar la base de datos desde la web en este trabajo de título, se utilizó la librería Selenium en Python.

3.1.2. Selenium

Selenium dispone de tres herramientas principales enfocadas al *scraping* de datos: Selenium WebDriver, Selenium IDE y Selenium Grid. En particular Selenium WebDriver permite controlar el navegador web tal como lo haría un usuario real. Esto quiere decir que todas las instrucciones dadas a través del código se podrán visualizar en pantalla a través del navegador determinado [24].

La librería Selenium en Python dispone de varios módulos y clases para la correcta ejecución del *scraping de datos*. Selenium cuenta con documentación para todos sus módulos, funciones y clases [25], a partir de la cual se detallan algunos de los módulos relevantes para este trabajo de título:

- `WebDriverWait`: se utiliza para generar tiempos máximos de espera en los cuales puede ser ejecutada una orden. Indica ignorar todas las instancias de *“NotFoundException”* que son encontradas en la ejecución determinadas por la condición *until*. Cuando se cumple la condición ejecuta la cadena de indicaciones siguientes.
- `expected_conditions`: determina condiciones para ejecutar acciones. Un ejemplo de estas condiciones es esperar hasta que un elemento determinado sea cliqueable. Esto significa que la siguiente acción (que podría ser clicar un botón específico) no se ejecutará hasta que encuentre que ese botón está habilitado para ser cliqueado.
- `By`: clase que se utiliza para localizar elementos dentro del código web. Por ejemplo, se puede realizar la selección por ID, por el nombre de la clase, etc.
- `find_elements`: función que se utiliza para localizar todos los elementos que respondan a la condición determinada. Retorna una lista de elementos.

A continuación, se muestran dos ejemplos que ilustran el uso de los métodos y clases antes descritos:

```
1 list_container = WebDriverWait(driver, 10).until(EC.presence_of_element_located((By.ID,
  ↪ "listadoOfertas")))
2
3 list_items = list_container.find_elements(By.CLASS_NAME, "result-box-container")
```

Código 3.1: Ejemplo 1. Uso de métodos y clases de Selenium. Elaboración propia.

La sección de Código 3.1, ejemplifica el uso de los cuatro métodos y clases descritos anteriormente. En esta sección de código, el objetivo es obtener una lista de resultados (`list_items`). Para esto, previamente se define la lista `list_container`, dentro de la cual se encuentra la lista de resultados buscada.

- `WebDriverWait(driver, 10).until(...)` espera como máximo 10 segundos. El tiempo de espera se reduce si se cumple la condición dada como parámetro dentro de *until*.
- `EC.presence_of_element_located()` es la condición determinada en la espera anterior, en este caso espera que el elemento dado como parámetro esté presente en el DOM

de la página. Es importante notar que EC es abreviatura determinada al importar `expected_conditions`

- *(By.ID, “listadoOfertas”)* está indicando que el elemento que se espera que esté presente es el ID denominado listadoOfertas.
- *find_elements(By.CLASS_NAME, “result-box-container”)*. Esta línea de código busca todos los elementos que pertenecen a la clase de nombre *result-box-container*, dentro de la lista guardada previamente denominada *list_container*.

3.2. Natural Language Processing (NLP)

En esta sección se revisan conceptos relevantes respecto al Procesamiento de Lenguaje Natural. En primera instancia se abordan temas relacionados a los modelos disponibles y sus diferencias. Para terminar la sección se revisan algunas herramientas de NLP aplicadas a la Minería de Textos.

El Procesamiento de Lenguaje Natural (NLP por sus siglas en inglés), es la rama de la Inteligencia Artificial (IA) que trata de dar entendimiento a los computadores de texto y palabras habladas de la misma forma en que las entiende un ser humano [26].

Según Jurafsky y Martin “la característica que logra distinguir al NLP de cualquier otro sistema de procesamiento de datos es el uso del conocimiento del lenguaje” [27, p. 2]. El conocimiento del lenguaje hace referencia a varios ámbitos de nuestro lenguaje:

1. Fonética y Fonología: los modelos de NLP deben ser capaces de entender y conocer los sonidos y pronunciaciones propios de cada idioma, para así poder generar audios o viceversa, entender audios y traducirlos a textos.
2. Morfología: para poder distinguir, por ejemplo, entre el plural y singular de una palabra, los modelos de NLP deben tener conocimiento respecto al significado de cada componente de las palabras.
3. Sintaxis: la frase: “*Chile en estudiando estoy*”, no tiene significado directo. Usando las mismas palabras se forma la frase “*estoy estudiando en Chile*”, la cual sí puede ser interpretada directamente. Para poder cumplir con los requerimientos actuales de un modelo de NLP, estos deben ser capaces de ordenar las palabras para dar respuestas coherentes.
4. Semántica: la semántica es el conocimiento del significado de las palabras. Los modelos de NLP deben ser capaces de entender a qué se refiere cada palabra utilizada, por ejemplo, al responder preguntas. ¿Qué idiomas se hablan en Sudamérica?. Para responder correctamente el modelo debe entender el significado de Sudamérica, qué países lo componen, etc.
5. Pragmática: la pragmática del lenguaje considera el contexto y la intención comunicativa que hay en un mensaje. Así, un modelo de NLP debe ser capaz de discernir, por ejemplo, entre una orden (abre la puerta), una afirmación (la puerta está abierta) y una pregunta (¿la puerta está abierta?).

- Discurso: para hacer posible herramientas como el conocido Chat GPT, el cual es capaz de obtener contexto a lo largo de toda la conversación, es necesario que el modelo sepa interpretar cómo se relacionan las distintas interacciones. En el ejemplo mostrado en la Figura 3.3, se puede ver que el modelo debe tener conocimiento discursivo para entender a qué se refiere la segunda pregunta.

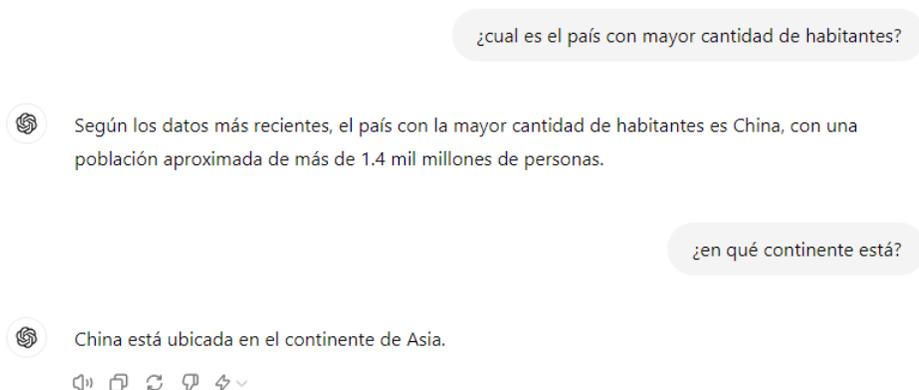


Figura 3.3: Ejemplo Chat GPT, conocimiento discursivo.

Todos estos puntos se encuentran descritos y ejemplificados con mayor detalle en el capítulo 1 de la segunda edición “Speech and Language Processing” de Daniel Jurafsky y James H. Martin [27].

Como resultado de este conocimiento, los modelos de NLP tienen variados usos en la actualidad. Algunos de los usos más conocidos se listan a continuación [26]:

- Clasificación de textos. Esta aplicación se basa en el análisis de textos con el objetivo de etiquetarlos según la necesidad. Algunos ejemplos de este tipo de uso pueden ser la detección de correos basura o el análisis de sentimientos en textos de redes sociales. Para ambas aplicaciones el modelo debe ser capaz de distinguir la intencionalidad de los mensajes y no sólo el significado literal de las palabras. Para este trabajo de título será la herramienta utilizada.
- Traducción automática. La traducción automática debe lograr transmitir el mensaje considerando el contexto y la intención de este. Google Translate es uno de los mejores y más masificados ejemplos [26].
- Agentes virtuales y *chatbots*. Alexa y Siri son dos tipos de agentes virtuales que utilizan el reconocimiento de voz para poder dar respuesta a las distintas solicitudes del usuario. Por otro lado, los *chatbots*, como Chat GPT, utilizan la misma metodología pero para entradas de tipo texto [26].

3.2.1. Large Language Models (LLM)

Los Large Language Models (LLM por sus siglas en inglés) son modelos preentrenados con grandes cantidades de datos de texto, de los cuales obtienen el conocimiento del lenguaje necesario para cumplir con los objetivos del NLP [28].

La arquitectura más usada para los LLM es la llamada “*transformer*” [29]. Esta arquitectura introduce el mecanismo de auto-atención para el entrenamiento de modelos. La auto-atención puede ser entendida como una forma de hacer que el modelo adquiriera conocimiento de las palabras considerando el contexto en el que están insertas, aprendiendo de la relación que tiene una palabra con otras en grandes extensiones de texto. En este mecanismo cada *token* de entrada se autoevalúa en relación a todos los otros *tokens* de la secuencia.

Para entender el mecanismo de autoatención se deben comprender 3 elementos claves:

1. *Key* (k): son todos vectores relacionados con los *tokens* que componen la cadena de entrada.
2. *Query* (q): es el vector correspondiente al *token* de interés.
3. *Value* (v): vector que contiene información respecto al contenido de los *tokens*. Se utiliza para ponderar la salida final.

En la Figura 3.4, se observa cómo funciona el mecanismo de autoatención para el cálculo del tercer *token* (a_3). La entrada al mecanismo es una secuencia de *tokens* a la que previamente se le aplicó un proceso de *embedding* para poder transformar palabras a vectores numéricos. En este caso, la secuencia de entrada es $X = [x_1, x_2, x_3]$, donde cada x_i corresponde a un *token*. A continuación, se listan y explican las etapas del mecanismo para el cálculo de a_3 :

1. Se calculan los vectores k , q y v para cada *token*. Estos se calculan mediante la multiplicación del vector de *embeddings* y las matrices de pesos autogeneradas de forma aleatoria, dependiendo del modelo de iniciación seleccionado.
2. El vector q correspondiente a x_3 se evalúa con los vectores k de cada *token* de la cadena, incluyendo el propio.
3. Los resultados pueden ser valores dentro del conjunto $\{-\infty, +\infty\}$ por lo que para escalar los resultados estos se dividen por d_k , que, típicamente, corresponde a la raíz cuadrada de la dimensión de k .
4. Las puntuaciones obtenidas se normalizan aplicando una función *softmax* para obtener finalmente los pesos de atención de cada *token* en relación al token de interés.

$$\alpha_{ij} = \frac{\exp(\text{score}(q_i, k_j))}{\sum_{k=1}^n \exp(\text{score}(q_i, k_k))} \quad (3.1)$$

donde α_{ij} es el peso de atención que indica la importancia del j -ésimo *token* en la secuencia al calcular la representación de i -ésimo *token*.

5. Cada vector α se pondera por su vector de valores (v) asociado.
6. Finalmente, se calcula la sumatoria de todos los vectores resultantes, para obtener a_3 (la nueva representación de x_3).

$$a_i = \sum_{j=1}^n \alpha_{ij} v_j \quad (3.2)$$

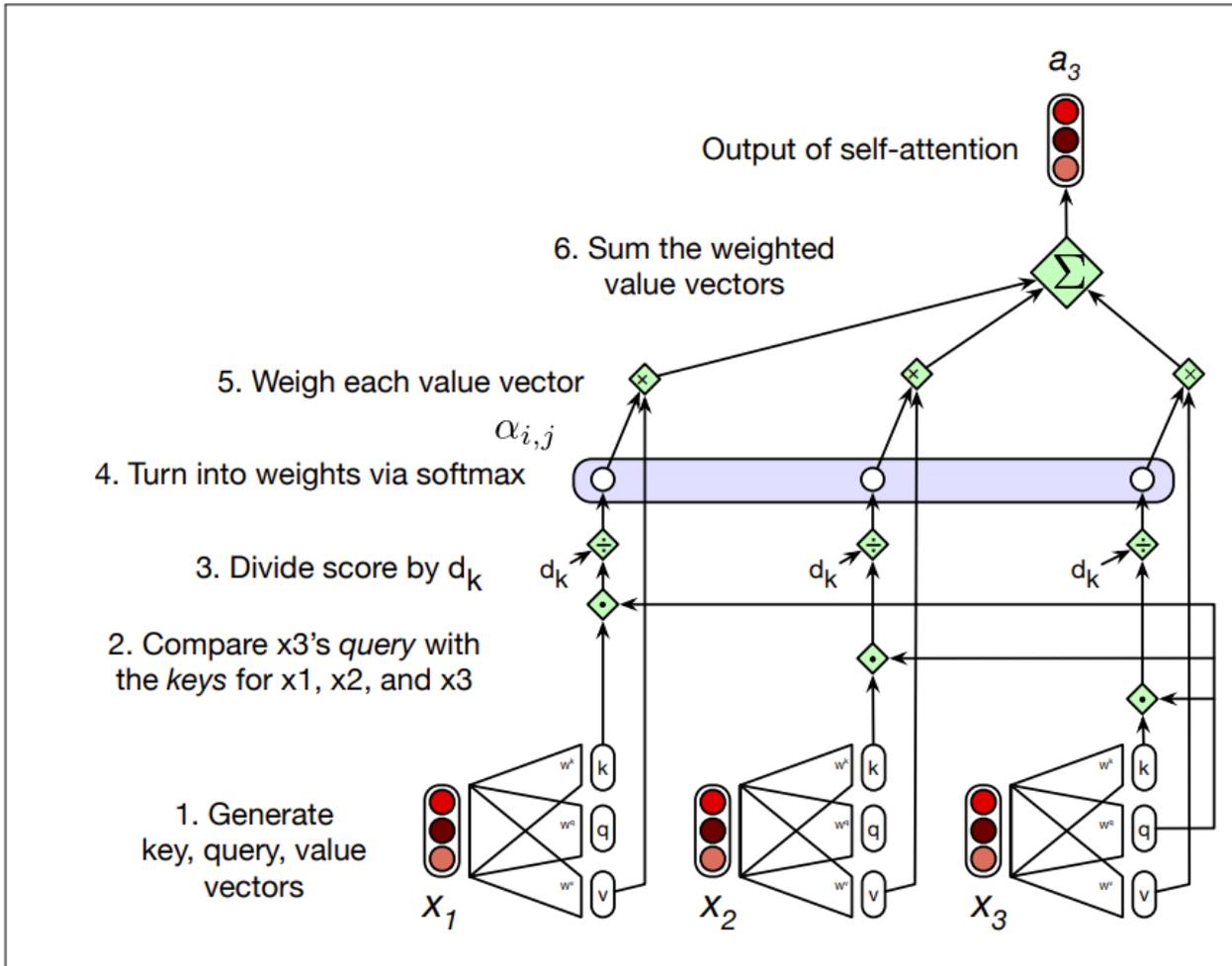


Figura 3.4: Cálculo del valor a_3 , el tercer elemento de una secuencia utilizando el mecanismo de auto-atención. Extraído de [29, p. 219].

El mecanismo de autoatención es la capa base para los modelos que utilizan la arquitectura de transformadores. Cada modelo utiliza la autoatención de distintas formas, dependiendo del objetivo.

La Figura 3.5 muestra 2 tipos de aplicaciones de las capas de autoatención. En la primera (a) llamada autoatención causal, la generación del *token* de salida depende de los *tokens* ingresados anteriormente. Por ejemplo, la generación de a_3 depende de a_1 , a_2 y a_3 . Por otro lado, en la autoatención bidireccional (b), se consideran todos los *tokens* al mismo tiempo.

Toda la teoría referida al mecanismo de autoatención fue extraída del libro *Speech and Language Processing*, tercera edición, de Jurafky D. y Martin J [29].

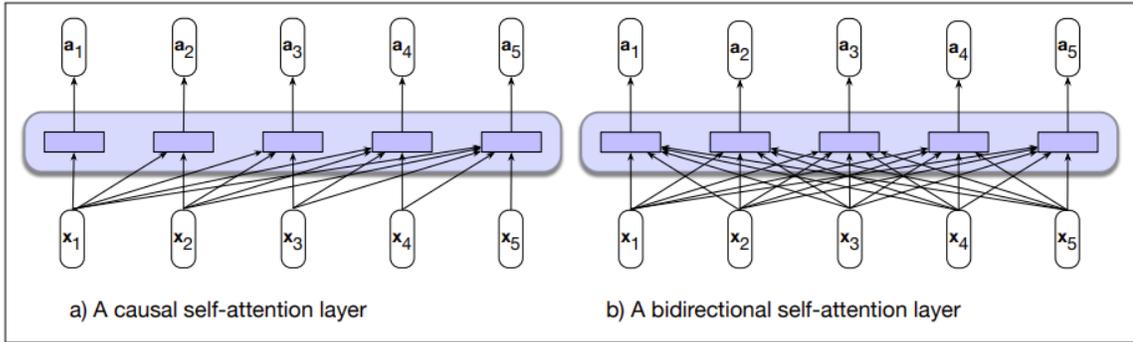


Figura 3.5: Capa de autoatención casual (a) y capa de autoatención bidireccional (b). Extraído de [29, p. 243].

A continuación, serán revisados dos modelos con base en esta arquitectura.

3.2.1.1. GPT-4o

GPT-4o (*Generative Pre-trained Transformer* en inglés), es un modelo basado en la arquitectura de los transformadores, pre-entrenado para predecir el siguiente *token* de un documento (no supervisado) usando dos fuentes de datos: la data pública de internet y data adquirida desde proveedores externos. El proceso de *fine-tuning* fue realizado usando *Reinforcement Learning from Human Feedback (RLHF)* [30]. En el mismo documento, OpenAI explica que se reserva otras características del modelo.

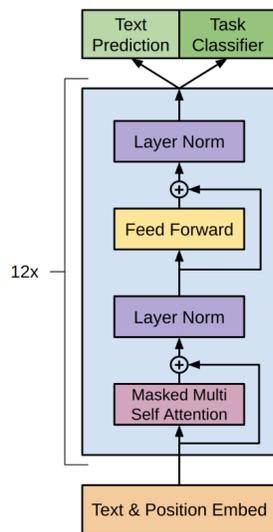


Figura 3.6: Arquitectura de transformador usada en GPT. Extraído de [31, p. 4].

La Figura 3.6 muestra la arquitectura

La salida de la capa de transformadores es ingresada al entrenamiento fino, dependiendo de la tarea objetivo [31].

utilizada para el modelo GPT base. La entrada a las capas es una cadena de texto embebido. A partir de ahí comienza el mecanismo de atención utilizado.

Cada capa contiene una subcapa de autoatención, considerando enmascarar las palabras futuras de la secuencia. Continúa con una subcapa de normalización. Luego, se incluye una capa de prealimentación (*feed forward*), que procesa cada posición de la secuencia de forma independiente, para así terminar con una segunda subcapa de normalización [31].

En la Figura 3.6 se observan 12 capas, ya que hace referencia al modelo GPT base. El modelo GPT-3 cuenta con 96 capas [32]. Para el modelo GPT-4, hasta el momento OpenAI se ha reservado los detalles de arquitectura.

3.2.1.2. BART

BART (*Bidirectional and Auto-Regressive Transformers* en inglés) es otro modelo de lenguaje basado en la arquitectura de transformadores. BART combina las estrategias de transformadores bidireccionales, con transformadores auto-regresivos. Esto es logrado utilizando un codificador bidireccional y un decodificador auto-regresivo [33].

El modelo BART puede ser entendido como una combinación de las estrategias de transformadores usadas en BERT y en GPT. La Figura 3.7, muestra una comparación de ambos modelos. En el lado izquierdo, se observa la estrategia de transformador bidireccional, en la que son enmascarados algunos *tokens* de la secuencia de forma aleatoria, buscando como salida una respuesta a los *tokens* ocultos. En el lado derecho, se observa el entrenamiento auto-regresivo, puesto que la sección enmascarada de la cadena de entrada son todos los *tokens* futuros al analizado. Se espera como respuesta una predicción al siguiente *token* en la cadena. Finalmente, para el modelo BART, estas dos estrategias de entrenamiento son combinadas, aplicando la primera en el codificador y la segunda en el decodificador del modelo [33].



Figura 3.7: Estrategias de transformadores. Derecha: estrategia usada en BERT. Izquierda: Estrategia usada en GPT. Extraído de [33, p. 2]

Finalmente, es importante mencionar que *bart-large-mnli* ha demostrado ser efectivo en tareas de generación de texto, traducción y comprensión del lenguaje natural [33], obteniendo un destacable rendimiento de 84.3% en *accuracy* en tareas de clasificación (MNLI).

3.2.1.3. Clasificación *Zero-Shot*

La clasificación *Zero-Shot*, es un tipo de tarea en el procesamiento de lenguaje natural, en la que un modelo es entrenado con conjunto de datos etiquetados, pero que luego es capaz de clasificar nuevos ejemplos con nuevas etiquetas no conocidas [34].

La Figura 3.8, muestra el proceso en particular para el modelo *bart-large-mnli*. Este modelo recibe un texto de entrada (texto que será clasificado) y una lista de etiquetas posibles. La salida del modelo consiste en una lista de probabilidades asociadas a cada etiqueta, donde la mayor probabilidad indicará la etiqueta seleccionada por el modelo.



Figura 3.8: Flujo de clasificación Zero-Shot (*bart-large-mnli*). Extraído de [34].

Para otros modelos de clasificación Zero-Shot, la salida puede variar en formato, pero en general la estructura es la misma.

3.2.2. Herramientas utilizadas para procesamiento de textos

En esta sección se revisan conceptos relevantes respecto al Procesamiento de Lenguaje Natural. En específico esta memoria recurre a las herramientas de NLP aplicadas a la Minería de Textos.

El Procesamiento de Lenguaje Natural (NLP por sus siglas en inglés), es la rama de la Inteligencia Artificial (IA) que trata de dar entendimiento a los computadores de texto y palabras habladas de la misma forma en que las entiende un ser humano [26]. Por otro lado, el objetivo de la Minería de Textos es el análisis y la extracción de información valiosa a partir de grandes cantidades de datos de texto no estructurado. Estas disciplinas convergen para ofrecer un enfoque integral en la exploración y comprensión de documentos.

3.2.2.1. Tokenización

La “tokenización” es la separación del documento en piezas unitarias. Cada una de estas piezas recibe el nombre de *token*. En ocasiones, al realizar este proceso, se eliminan símbolos de puntuación y otros [35].

Existe también el término “tipo”, que hace referencia al conjunto de *tokens* que tienen la misma secuencia de caracteres. Es importante definir “término”, un “tipo” que está incluido en el vector de características del documento.

Dentro de los métodos especializados para realizar la “tokenización” se encuentra el dispuesto por spaCy y BERT Tokenizer.

3.2.2.2. spaCy

SpaCy es una librería especializada en Procesamiento de Lenguaje Natural, gratuita y de código abierto en Python [36]. Está construida para ser capaz de procesar documentos, “tokenizar”, identificar palabras vacías, puntuación, etc. A partir de ella, se pueden procesar

grandes cantidades de texto con el objetivo de extraer información, alimentar sistemas que entiendan el lenguaje natural o preprocesar texto para ser trabajado con herramientas de aprendizaje profundo.

3.2.2.3. Análisis de Componente Principal (PCA)

La idea central del Analisis de Componente Principal, es reducir la dimensionalidad de un conjunto de datos con gran cantidad de variables correlacionadas [37]. Este objetivo se logra transformando el conjunto de datos a un nuevo set de variables no correlacionadas, ordenadas desde la variable que retiene más variabilidad a la que retiene menor variabilidad.

A continuación, se revisa la definición formal de PCA. Se supone un vector x con p variables aleatorias. Además, la estructura de las co-varianzas y las correlaciones entre las p variables son de interés. Cuando p es grande, resulta poco útil y complejo visualizar las correlaciones y co-varianzas o generar visualizaciones de los datos. Dado este problema, la alternativa propuesta por PCA es considerar sólo las variables p que contienen la mayor cantidad de información relacionada a la co-varianza y correlaciones [37].

PCA se concentra en las varianzas de los datos. El primer paso es buscar la función lineal $\alpha'_1 x$ con máxima varianza, donde α'_1 es un vector de p constantes transpuesto. Así,

$$\alpha'_1 x = \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1p}x_p = \sum_{j=1}^p \alpha_{1j}x_j. \quad (3.3)$$

A continuación, se busca la función lineal $\alpha'_2 x$ con la máxima varianza, pero que no esté correlacionada con α'_1 . Así, se continúa iterando hasta encontrar los m componentes principales. Es importante mencionar que m puede ser igual a p , pero no es útil ni esperable puesto que el objetivo de PCA, es representar la mayor cantidad de información en los primeros componentes principales.

3.3. Clustering

El *clustering* o agrupación, es un tipo de aprendizaje no supervisado, ya que las instancias u observaciones en el conjunto de datos no vienen etiquetadas por lo que el modelo no puede verificar o comprobar que las predicciones que realiza son correctas. Los métodos de *clustering* “aprenden” a través de la observación de las muestras y las agrupan según similitud de características [38].

Existen distintos tipos de *clustering*, clasificados según la forma de agrupar datos. Los que se describen en este documento son métodos de agrupación basados en centroides y agrupación jerárquica.

3.3.1. Agrupación basada en centroides

Es el método de agrupación más común. Se basa en la configuración de centroides, a partir de los cuales se busca minimizar la sumatoria de las distancias de los puntos pertenecientes al clúster con el centroide. El mejor ejemplo para este método de agrupación es el algoritmo K-Means.

3.3.1.1. K-Means

El algoritmo K-Means, es el método de agrupación más conocido y utilizado. El objetivo es minimizar la distancia euclidiana promedio entre cada punto del clúster y su centro [39]. El centro del clúster se define como el promedio, o centroide del grupo, usando la Ecuación 3.4, donde ω es el conjunto de puntos \vec{x} perteneciente al clúster.

$$\vec{\mu} = \frac{1}{|\omega|} \sum_{x \in \omega} \vec{x} \quad (3.4)$$

Para evaluar qué tan bien los centroides están representando a los puntos se utiliza la Suma Residual de Cuadrados (RSS por sus siglas en inglés), para esto se calcula el RSS para cada cluster (Ecuación 3.5), y luego se hace la suma de todos los resultados (Ecuación 3.6).

$$RSS_k = \sum_{\vec{x} \in \omega_k} |\vec{x} - \mu(\vec{\omega}_k)|^2 \quad (3.5)$$

$$RSS = \sum_{k=1}^K RSS_k \quad (3.6)$$

Algoritmo

1. Definir K centroides aleatoriamente.
2. Asignar cada punto al centroide más cercano según la distancia euclidiana.
3. Recalcular los nuevos centroides.
4. Iterar desde el paso 2 hasta que se cumpla la condición de parada.

Las condiciones que proponen Manning, C. et. al, son las siguientes [39]:

- Determinar una cantidad fija de iteraciones. Esta opción puede mejorar el rendimiento computacional y reducir tiempos pero puede no generar buenos resultados, puesto que no se considera si el algoritmo llegó al punto óptimo o no.
- Detener el algoritmo cuando la posición de los centroides no cambia entre iteraciones. Esta opción en general obtiene buenos resultados, exceptuando cuando el algoritmo cae en un mínimo local. Además, puede generar que el tiempo que toma correr el algoritmo sea extremadamente largo.
- Detener el algoritmo cuando los puntos asignados que son asignados a cada centroide no cambian entre iteraciones. Esta opción es equivalente a la anterior.
- Detener el algoritmo cuando el valor de RSS cae por debajo de un umbral determinado. En general, esta opción debe combinarse con la determinación de una cierta cantidad de iteraciones, puesto que puede no terminar nunca.
- Terminar el algoritmo cuando la disminución de RSS entre una iteración y otra es menor a un umbral determinado, es decir, la mejora en la calidad de los resultados del algoritmo entre una iteración y otra, es menor a este umbral.

Finalmente, en la Figura 3.9, se puede observar la evolución durante la aplicación del algoritmo K-Means para un set de datos. Al iniciar el algoritmo se configuran 2 centroides aleatoriamente, que se van reubicando en cada iteración, hasta obtener el resultado final en la novena iteración.

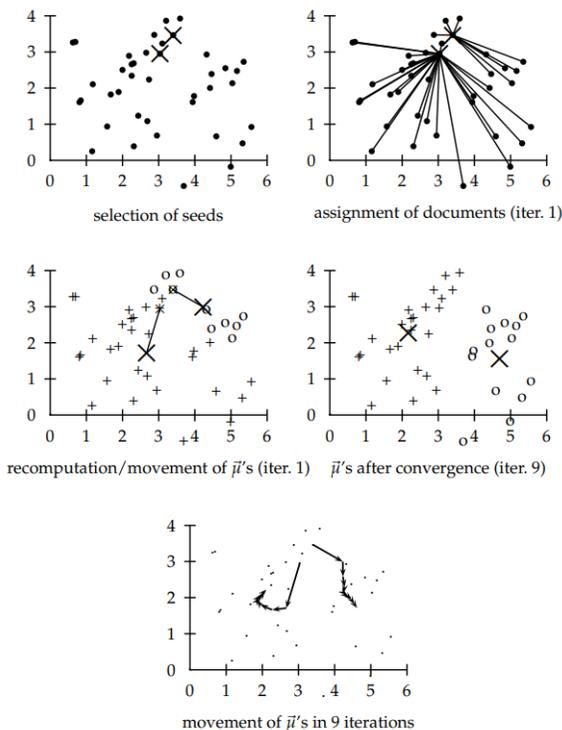


Figura 3.9: Ubicación de centroides según iteraciones. Extraído de [39, 362].

3.3.2. Agrupación basada en la estructura jerárquica

El algoritmo de agrupación jerárquico, es un método de agrupación basado en la estructura jerárquica de los datos [40]. A diferencia de otros métodos de agrupación que asignan cada punto de datos a un clúster específico, la agrupación jerárquica crea una estructura de clústers anidados o en forma de árbol. Existen dos tipos de agrupación jerárquica: Aglomerativa y Divisiva.

El tipo Aglomerativo comienza considerando cada uno de los elementos como un clúster individual, y de forma iterativa los combina hasta obtener un único clúster. La Figura 3.10, vista de izquierda a derecha, muestra un ejemplo de la estrategia de agrupación aglomerativa para los objetos a , b , c , d y e . AGNES (*Agglomerative Nesting*) es un método de agrupación aglomerativa [40].

La estrategia divisiva comienza considerando a todos los elementos dentro de un único clúster, llamado raíz. De forma recursiva este clúster es dividido en clústeres más pequeños, hasta obtener un clúster por cada objeto. La Figura 3.10, vista de derecha a izquierda, muestra un ejemplo de este tipo de estrategia de agrupación. DIANA (*Divisive Analysis*) es un método de agrupación divisiva [40].

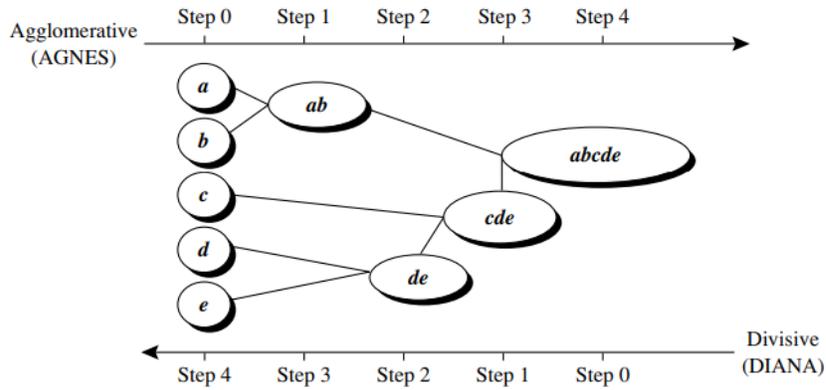


Figura 3.10: Estrategias de agrupación aglomerativa y divisiva. Extraído de [40, p. 460].

En el método AGNES, para cada etapa es evaluado el criterio seleccionado, por ejemplo, la distancia Euclídeana. Cada clúster es representado por todos los elementos que pertenecen a él. En cada nivel se unifican los clústeres con la distancia mínima entre ellos, hasta quedar con un único clúster final. El método DIANA funciona de forma inversa, considerando la distancia máxima entre los vecinos cercanos a un punto para separarlos en clústeres diferentes. Se detiene cuando se obtiene un clúster por cada elemento [40].

La Figura 3.11 muestra un dendrograma, que es la forma más típica de visualizar los resultados de agrupación jerárquica. Según esta visualización, se puede escoger la cantidad de clústeres deseados, considerando las distancias entre ellos [40].

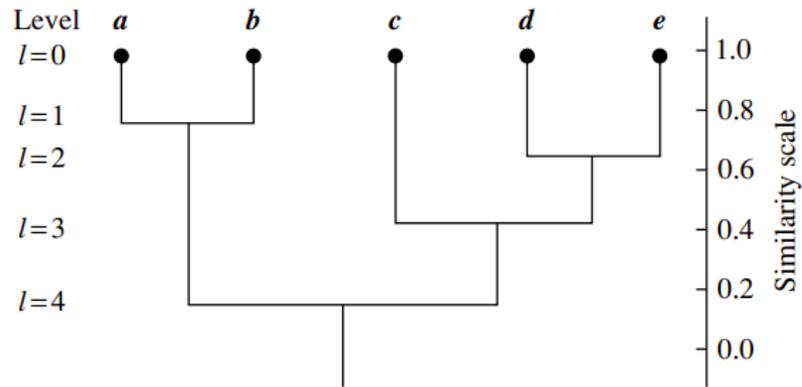


Figura 3.11: Dendrograma típicamente usado para representar los resultados de agrupación jerárquica. Extraído de [40, p. 460].

Capítulo 4

Etapas y Metodología del Proyecto

En este capítulo se presenta la metodología llevada a cabo en el presente trabajo de título.

El objetivo principal de este trabajo de título es formular una propuesta para realizar mejoras en el currículum y organización de los cursos del Departamento de Ingeniería Eléctrica a partir de análisis realizados con el uso de herramientas de Inteligencia Artificial.

El análisis mencionado previamente se lleva a cabo utilizando una gran cantidad de ofertas de trabajo publicadas en la web dirigidas a egresados de la carrera Ingeniería Civil Eléctrica en Chile. Para esto se etiquetan las ofertas de empleo utilizando *prompts* en GPT, según la lista de competencias generada para este objetivo.

Este trabajo de título se organiza en cinco etapas:

1. Recolección de datos.
2. Pre-procesamiento y limpieza de datos.
3. Etiquetado de la base de datos.
4. Construcción de la propuesta.
5. Consolidación de la propuesta.

En la Figura 4.1 se presenta el flujograma de las etapas y metodología utilizada.

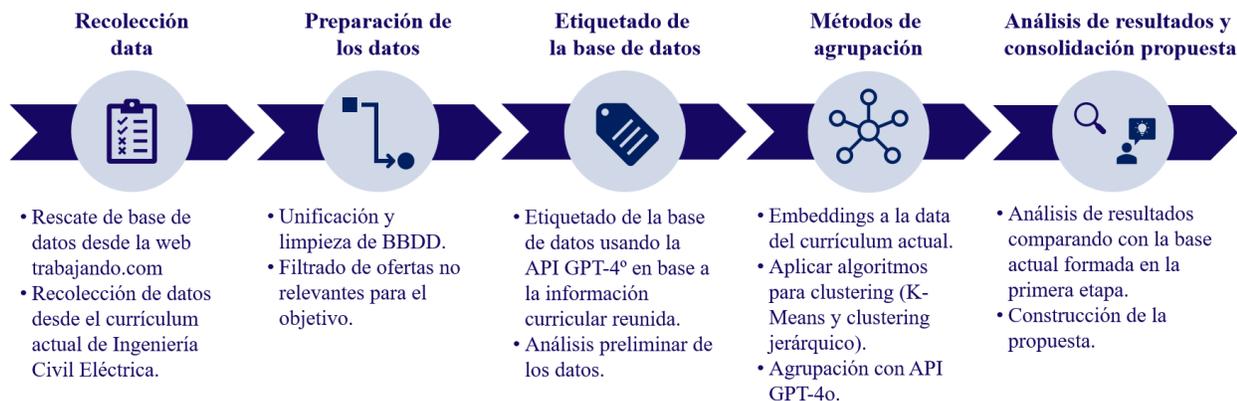


Figura 4.1: Etapas y Metodología trabajo de título.

4.1. Recolección de datos

La recolección de datos para alcanzar los objetivos de este trabajo de título se aborda desde dos perspectivas diferentes. Por un lado se realiza la recolección de ofertas de empleo publicadas en la web de trabajando.com para construir la base de datos central. Adicionalmente, se hace una revisión de los cursos obligatorios de la malla de Ingeniería Civil Eléctrica relacionados al plan de formación en innovación.

En esta sección se detallan los procedimientos de recolección de datos desde ambas perspectivas, comenzando con el análisis de plan curricular actual de Ingeniería Civil Eléctrica.

4.1.1. Análisis del currículum de Ingeniería Civil Eléctrica

La revisión del plan de formación en innovación actual del Departamento de Ingeniería Eléctrica de la Universidad de Chile, tiene como objetivo obtener un mapa de competencias, subcompetencias y resultados de aprendizaje, a partir de los cuales se realiza la clasificación para el análisis final.

Como este trabajo de título se encuentra enfocado en la formación en innovación del plan curricular de la carrera antes mencionada, se realiza el análisis de los cursos enfocados en esta competencia:

1. EL3105: Seminario de Ingeniería Eléctrica e Innovación Tecnológica.
2. EL4204: Formulación y Evaluación de Proyectos.
3. EL6101: Taller de Proyectos Tecnológicos.
4. EL6201: Taller de Emprendimiento e Innovación.

Se excluyen de análisis los cursos relacionados a innovación de la fase inicial de la carrera (Plan Común), y los cursos electivos disponibles.

Se revisa el programa de cada uno de los cursos mencionados en la lista, a partir de los cuales se identifican las competencias específicas y genéricas que se pretenden abordar en cada uno de los cursos, y que responden al perfil de egreso de la carrera.

Este análisis se realiza en una matriz donde cada columna corresponde a una competencia genérica o específica y las filas corresponden a los 4 cursos mencionados. Se indica con un 1 la presencia de la competencia en el programa y, con un 0 la ausencia.

De las competencias específicas, se consideraran relevantes para este trabajo de título las indicadas con los números 2, 3, 4, 5 y 6. De las competencias genéricas, se consideraran las competencias 1, 4, 6 y 7. A partir de estas competencias, se realiza un mapeo de los resultados de aprendizaje que responden a ellas. El resultado se organiza en dos redes tripartitas que relacionan: competencia, código del curso y resultado de aprendizaje. La diferencia entre ellas es que la primera considera sólo competencias específicas, y la segunda, considera sólo competencias genéricas.

A partir de estos resultados de aprendizaje, se le solicita a Chat GPT-4o que identifique competencias y sub-competencias relacionadas al área de innovación. La indicación incluye la especificación de que estos resultados de aprendizaje pueden ser agrupados bajo grupos de competencias, o desglosados. El uso de la aplicación de chat antes que la API en esta etapa, está motivado por la utilidad de la memoria para introducir contexto en la conversación. Previo al ingreso de la solicitud, se da contexto mediante el documento *Big Data in Medical Informatics: Improving Education Through Visual Analytics* [9], preguntando si tiene conocimiento de la publicación mencionada. La indicación dada para obtener los resultados es la siguiente:

Tengo como punto de partida 37 resultados de aprendizaje distintos del plan de formación en innovación de la carrera Ingeniería Civil Eléctrica. Quiero agrupar estos 37 resultados de aprendizaje en competencias y subcompetencias para generar una estructura similar a la publicación, ¿me ayudas?

A continuación, se facilitan los resultados de aprendizaje obtenidos de los programas de los cursos estudiados. Es importante mencionar que los resultados de aprendizaje pueden aparecer relacionados a más de una competencia específica o genérica, por lo que para las clasificaciones cada uno se cuenta una sola vez (a pesar de mantener la información de las relaciones).

Adicionalmente, se le solicita a la herramienta de Inteligencia Artificial la creación de etiquetas que representen a cada uno de los 37 resultados de aprendizaje. La indicación entrada es la siguiente:

Finalmente, necesito que para cada resultado de aprendizaje generes etiquetas que representen el detalle de cada uno. La idea de esto es luego clasificar con la API de GPT4o varias ofertas de trabajo identificando si calzan con ese resultado de aprendizaje, por lo que las etiquetas deben ser autoexplicativas y no muy largas.

Este resultado es guardado en una tabla que relaciona el resultado de aprendizaje con la etiqueta.

Finalmente, y a partir de los resultados entregados en la conversación, se organizan las competencias y sub-competencias bajo tres competencias mayores. Cada una de estas apunta a un aspecto de los distintos saberes (saber, saber hacer, saber ser). El objetivo es generar una organización estructural similar a la propuesta en el documento citado en [9], pero a partir del análisis entregado por la inteligencia de OpenAI y aplicado a la carrera de Ingeniería Civil Eléctrica de la Universidad de Chile.

Para estos resultados, se generan visualizaciones utilizando redes multipartitas, con el objetivo de hacer visuales las conexiones de cada resultado de aprendizaje y los distintos niveles de organización.

4.1.2. Recolección de ofertas de trabajo en la web

Como fue mencionado previamente, la base de datos de este trabajo se construyen a partir de ofertas de trabajo publicadas en la web, ya que no existen bases de datos útiles para el

objetivo de este trabajo de título.

El portal de empleos seleccionado para extraer la información es *trabajando.com*. El formato de página utilizado en este portal resulta particularmente útil para la aplicación de técnicas de *web scraping*, puesto que el modelo DOM es claro.

4.1.2.1. Organización de la web *trabajando.com*

La configuración de *trabajando.com* es bastante clara y, en particular, las ofertas se encuentran bien seccionadas, lo que facilita el poder acceder a los datos y estructurar la base.

La Figura 4.2, muestra la configuración visual de la página web. Se puede observar que la primera sección, marcada en verde, presenta el título de la oferta, el logo de la empresa que realiza la publicación, el día de publicación y el nombre de la empresa que ofrece el empleo. En la segunda sección, marcada en color azul, se incluye toda la información relacionada al detalle de la oferta laboral. En esta sección las empresas suelen incluir los beneficios, labores y requisitos para el cargo. Finalmente, la tercera sección (rojo) se titula siempre “Perfil deseado”, donde se incluye el perfil que se busca para el cargo en cuestión. Muchas veces esta sección repite información que ya fue presentada en la segunda sección, como en el ejemplo de la figura.

Analizando la visualización del árbol DOM, esta división de secciones resulta clara para poder acceder a cada uno de los campos. La Figura 4.3 muestra el árbol DOM correspondiente a la sección de la página web mostrada en la Figura 4.2.

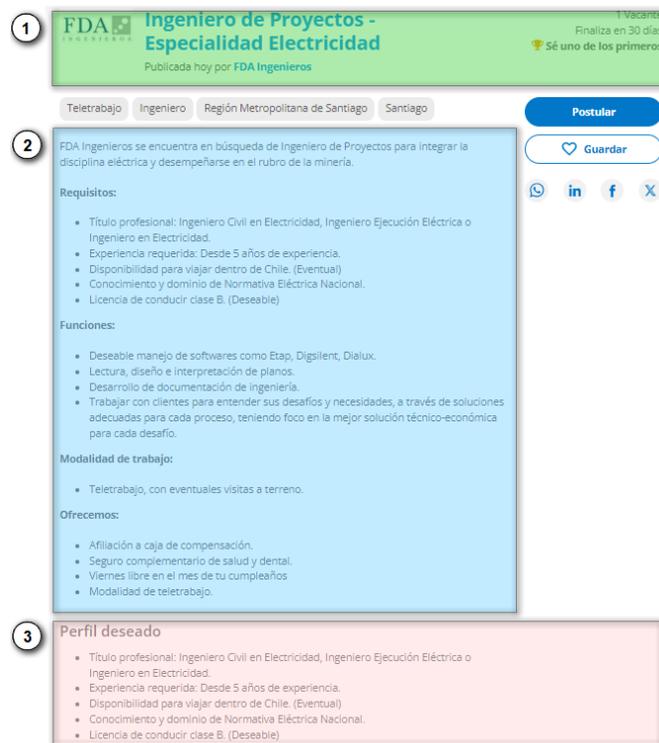


Figura 4.2: Organización de la oferta de empleo *trabajando.com*

En la Figura 4.3, se puede observar en color azul todos los componentes de la página que son de interés para la construcción de la base de datos. Se describen, según la numeración, a continuación:

1. h3: el componente indicado con h3 en el árbol DOM corresponde al título de la oferta, correspondiente al título indicado con letras azules en la sección 1 de la Figura 4.2.
2. text: es un componente de texto que contiene la frase “Publicado por”.
3. a: es el componente que contiene el nombre de la empresa que publica la oferta. Este componente tiene una url asociada, que direcciona al perfil de la empresa.
4. div: el componente div, etiquetado bajo el número 4, corresponde a una sección (con varias subdivisiones) que contiene toda la información correspondiente al detalle de la oferta laboral. Recordando la Figura 4.2, es la sección indicaba con el número 2 (azul).
5. div: el componente div, marcado con el número 5, corresponde a la sección “Perfil deseado” (tercera sección (roja) de la Figura 4.2).

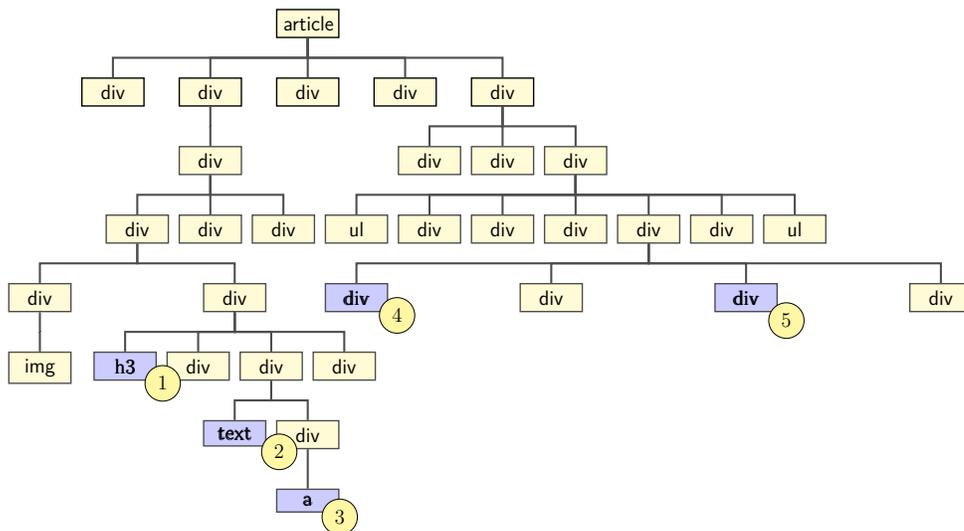


Figura 4.3: Árbol DOM elementos de interés en *trabajando.com*

4.1.2.2. *Web scraping*

Considerando la organización de la página web de *trabajando*, detallada en la subsección anterior de este documento, se realiza el proceso de recolección de datos mediante técnicas de *web scraping*. El trabajo de código se desarrolla en Python, usando las librerías disponibilizadas por Selenium.

Es importante mencionar que el uso de la automatización para recuperar la base de datos, se aplica desde la búsqueda inicial desde el navegador de *trabajando.com* hasta la revisión de cada uno de los resultados de esta búsqueda.

La selección de términos para la búsqueda de ofertas, se realiza considerando las 5 áreas de la carrera de Ingeniería Civil Eléctrica (Inteligencia Artificial, Energía, Telecomunicaciones, Control de Sistemas e Instrumentación). Así, los textos de búsqueda son:

1. Ingeniero Eléctrico.
2. Ingeniero Electrónico.
3. Ingeniero Telecomunicaciones.
4. Ingeniero de Proyectos Eléctricos.
5. Analista de datos.
6. Ingeniero TI.
7. Inteligencia Artificial.

El formato de salida de cada una de estas ejecuciones del código es un archivo csv con las ofertas recolectadas. Las columnas de este csv corresponden a la información extraída de la página web. A continuación, se enumeran las columnas de la base de datos construida:

1. ID
2. Título
3. Empresa
4. Contenido
5. Perfil

4.2. Preparación de los datos

Tras ejecutar el código de *web scraping*, se unifica la base de datos en un sólo archivo csv para ser procesado.

Es importante considerar que al realizar búsquedas que tengan relación unas con otras, se pueden obtener resultados repetidos en las ofertas de empleo entregadas por el buscador de la página web. Es por esto que en esta etapa, también se implementó la limpieza de datos duplicados y de emoticones en el conjunto de datos.

Finalmente, tras unificar el conjunto de datos en un sólo archivo csv, y realizar la limpieza de los datos recolectados, se procede a eliminar las ofertas de trabajo que no son relevantes para esta investigación, aplicando herramientas de Inteligencia Artificial. Este último paso se encuentra motivado por la inspección de las ofertas obtenidas. Es común que al buscar información en la web, los primeros resultados sean más coherentes con la búsqueda que los últimos que aparecen (menos prioritarios). El objetivo es eliminar los resultados de ofertas de empleo entregadas por *trabajando.com* que no tienen importancia para esta investigación.

4.2.1. Unificación y limpieza de la base de datos

Para llevar a cabo esta primera etapa de la preparación de los datos se utiliza la librería *Pandas* de *Python*. Se utiliza la función *read_csv* para leer los archivos generados en cada una de las búsquedas, los que son guardados en *dataframes* separados. Los nombre estos *drataframes* son guardado en una lista.

Se crea el *dataframe* final en el cual se implementa la estructura de la base de datos utilizada y, en este, se anexa la totalidad de ofertas de empleo recuperadas de la web.

Se identifican dos problemas en los datos recuperados. Por un lado, se tiene en cuenta la posibilidad de haber rescatado ofertas de empleo duplicadas, ya que hay búsquedas que pueden haber arrojado resultados comunes. Además, varias de las ofertas contienen “emoticones” o caracteres especiales, lo que puede causar problemas al momento de ingresarlos a modelos de procesamiento de lenguaje natural.

Para resolver la duplicidad de datos en la base, se utilizó el método *duplicated* de la librería Pandas. Este método identifica registros duplicados automáticamente dentro de un *dataframe*. Además, para controlar y tener conocimiento respecto a la cantidad de duplicados se obtienen los valores mediante el método *sum*.

Una vez que es identificado el total de registro duplicados, se procede con la eliminación de ellos, utilizando el método *drop_duplicates*.

Finalmente, para abordar el problema de los emoticones presentes en las ofertas de empleo, se construye la función *eliminate_emojis*. Esta función utiliza las cadenas de textos como tipo de dato determinado tanto para la entrada como para la salida. La función es aplicada a todo el conjunto de datos, modificando los registros en caso de existir emoticones en la oferta de empleo original.

4.2.2. Eliminación de registros no relevantes

Para seleccionar el mejor método para identificar ofertas de empleo no relevantes, se lleva a cabo un experimento, en el cual se compararan los resultados de dos métodos de clasificación. En el primero, se utiliza clasificación *Zero-Shot* con el modelo *bart-large-mnli*, y el segundo, utiliza *prompt engineering* mediante la API de GPT-4o.

Los datos seleccionados para esta prueba son los resultados correspondientes a la búsqueda “Inteligencia Artificial”. Este conjunto de datos se escoge porque contiene sólo 30 registros, haciendo más simple la tarea de clasificarlos de forma manual, para posteriormente evaluar el desempeño de las clasificaciones de los modelos utilizados.

Para ingresar los datos al modelo se unen en una sola cadena de texto los valores de las columnas “Título”, “Descripción” y “Perfil deseado”, ya que el primer modelo recibe como entrada una única cadena de texto a clasificar.

Las etiquetas utilizadas para el experimento son “Inteligencia Artificial” y “Otros”. En base a ellas se le indica a los modelos clasificar las ofertas de trabajo entregadas. Así, el objetivo del experimento consiste en identificar cuales ofertas obtenidas en la búsqueda corresponden efectivamente a ofertas relacionadas a Inteligencia Artificial.

Para evaluar el desempeño de los modelos, se utilizan las matrices de confusión y se obtienen las métricas de desempeño *accuracy*, *precision*, *recall* y *f1-score*, proporcionadas por la librería *scikit-learn* de Python.

4.2.2.1. Identificación de registros no relevantes: Clasificación Zero-Shot

Para realizar la clasificación Zero-Shot, se utiliza el modelo *facebook/bart-large-mnli*. Este modelo especializado en clasificación recibe como entrada una lista de etiquetas disponibles y una cadena de texto a clasificar. La salida del modelo es un diccionario con dos claves: “*labels*” y “*score*”. Así, se implementa un bucle para realizar una llamada al modelo por cada oferta de trabajo.

El resultado de la implementación en bucle es un *dataframe* con las columnas descritas a continuación:

- *oferta*: contiene las cadenas de texto clasificadas. Cada observación de *dataframe* corresponde a una oferta de trabajo.
- *labels*: guarda la etiqueta escogida por el modelo para la cadena de texto analizada. Esta corresponde a la opción con mayor probabilidad entregada por el modelo.
- *IA*: guarda la probabilidad dada por el modelo de que la cadena de texto analizada corresponda a la categoría “Inteligencia Artificial”.
- *OTR*: guarda la probabilidad dada por el modelo de que la cadena de texto analizada corresponda a la categoría “Otros”.

Se guardan los resultados en un archivo csv para ser procesados y analizados.

4.2.2.2. Identificación de registros no relevantes: API GPT

La clasificación con la API de GPT utiliza el modelo GPT 4o. Este modelo requiere de tres entradas para realizar la clasificación. La primera es el modelo de GPT que se desea utilizar en la llamada. La segunda y tercera entrada, corresponden a los mensajes que se le quieren entregar al modelo con el contexto e indicaciones de la tarea que debe realizar.

Se crea la función *clasificar_oferta*, dentro de la cual se realiza la llamada a la API de GPT. Esta función recibe como parámetros una oferta de empleo (cadena de texto) y una lista con las posibles categorías.

Para realizar la clasificación y limitar la respuesta del modelo dentro de los márgenes requeridos para obtener resultados, se ingresan dos cadenas de texto al modelo, bajo los roles *system* y *user*. Estas fueron:

1. Cadena de contexto (*system*): *Me ayudarás a clasificar las ofertas de trabajo decidiendo a cual categoría de la lista: {', '.join(categories)} corresponde la oferta. Además, para cada categoría añadirás la probabilidad de que la oferta corresponda a esa categoría. No usar el símbolo de doble comilla en la respuesta.*
2. Indicación o *prompt* (*user*): *Analiza {oferta} según el contexto dado. Entrega una lista de la forma [probabilidad de que pertenezca a Inteligencia Artificial, probabilidad de que pertenezca a Otros]. No entregues información adicional en la respuesta, y la probabilidad debe ser entre 0 y 1.*

Como se puede observar, ambas cadenas de texto de son funciones de Python, ya que reciben los parámetros *categories* y *oferta*, respectivamente.

La respuesta a la llamada a la API de GPT es siempre una cadena de texto, por lo que para obtener una lista con las probabilidades, como se le solicita al modelo, es necesario hacer la transformación a una lista mediante el separador “,”.

Esta función es aplicada en un bucle, iterando sobre cada una de las ofertas de empleo del conjunto de datos determinado.

Para conseguir guardar la información entregada por GPT en un conjunto de datos idéntico en estructura al descrito para la clasificación *Zero-Shot*, es necesario realizar una limpieza y análisis de los datos, puesto que sólo se tienen las probabilidades de que la oferta perteneciera a una etiqueta u otra.

Para esto se implementa la función *extract_number*, la cual recibe como entrada una cadena de texto, elimina todos los caracteres especiales dentro de ella y realiza la transformación a formato *float*.

Se implementa la función en todo el conjunto de datos para obtener las probabilidades en formato numérico y poder identificar la respuesta de clasificación definitiva del modelo.

Finalmente, se exportan los resultados a un archivo csv.

4.2.2.3. Evaluación de desempeño de los modelos

Los resultados de ambos modelos son guardados en archivos csv de respaldo. Las etiquetas definidas de forma manual para cada oferta también son guardadas en archivo csv. Así, para evaluar el desempeño de cada modelo, se importan los tres archivos con los resultados.

A partir de las etiquetas obtenidas se definen las siguientes variables:

1. *y_true*: contiene las etiquetas creadas a manualmente, es decir, el resultado esperado de clasificación.
2. *y_zero*: contiene las etiquetas asignadas por la clasificación Zero-Shot.
3. *y_gpt*: contiene las etiquetas asignadas por el modelo GPT 4o.

Con estas tres variables son construidas las matrices de confusión correspondientes, utilizando *ConfusionMatrixDisplay* de la librería sklearn. Se utiliza esta misma librería para obtener las métricas antes mencionadas.

4.2.2.4. Filtro final de la base de datos

Es importante mencionar que el filtro final es realizado al conjunto de datos completo, y no a cada uno de los resultados de búsqueda por separado. Esta decisión está motivada por el hecho de que al clasificar individualmente cada resultado, se pueden perder ofertas de empleo que quizás no califican en la etiqueta de la búsqueda asociada, pero sí en alguna otra.

En base al experimento antes descrito, se determina que el mejor modelo a utilizar para realizar el filtro de las ofertas calificadas como no relevantes para el objetivo de esta memoria, fue GPT 4o. Por lo que se utiliza un procedimiento similar al descrito anteriormente.

Las diferencias de implementación radican en la lista de posibles clasificaciones y la indicación ingresada al modelo. Estas diferencias se detallan a continuación.

La lista de etiquetas difiere de la utilizada en el experimento para Inteligencia Artificial, puesto que para realizar el filtro final se consideraran como etiquetas las ochos las búsquedas realizadas, más la categoría Otros. Así, la lista de etiquetas fue: Ingeniero Eléctrico, Ingeniero Electrónico, Ingeniero Telecomunicaciones, Ingeniero de Proyectos Eléctricos, Analista de datos, Ingeniero de datos, Ingeniero TI, Inteligencia Artificial y Otros.

La indicación ingresada al modelo en el experimento de clasificación entre dos etiquetas es bastante específica en cuanto al formato de salida de la respuesta. Según esto, es necesario modificar las cadenas de textos ingresadas para obtener el resultado en formato de lista, como es requerido. Las cadenas de texto ingresadas son las siguientes:

1. Cadena de contexto (*system*): *Me ayudarás a clasificar las ofertas de trabajo decidiendo a cual categoría de la lista: ({', 'join(categories)}) corresponde la oferta. Para eso calcularás la probabilidad de que la oferta corresponda a esa categoría. No usar el símbolo de doble comilla en la respuesta.*
2. Indicación o *prompt* (*user*): *Analiza {oferta} según el contexto dado. Entrega una lista de la probabilidad de que pertenezca a cada categoría en el orden [Ingeniero Electrico, Ingeniero Electronico, Ingeniero Telecomunicaciones, Ingeniero de Proyectos Electricos, Analista de datos, Ingeniero de datos, Ingeniero TI, Inteligencia Artificial, Otros]. No entregues información adicional en la respuesta. La probabilidad debe ser entre 0 y 1, y la suma de las 9 probabilidades debe ser 1.*

Los resultados son guardados en un conjunto de datos con la misma configuración anterior, donde las dos primeras columnas guardan la oferta analizada y la etiqueta con mayor probabilidad asignada, respectivamente. El resto de columnas corresponden a las probabilidades de las etiquetas, en el mismo orden nombradas previamente. Se exporta el resultado a un archivo csv.

Finalmente, con las clasificaciones entregadas por el modelo, se filtra la base datos construida quitando todas las ofertas de empleo que resultan clasificadas en la categoría “Otros”. Esto deja a disposición de este trabajo de título, una base con ofertas de trabajo relevantes para el objetivo definido.

4.3. Etiquetado de la base de datos

Tras realizar la limpieza de la base de datos, incluyendo el filtro de ofertas no relevantes para el objetivo de este trabajo de título, el siguiente paso es el etiquetado de esta.

El objetivo de esta etapa es identificar qué resultados de aprendizaje de la lista obtenida en la primera etapa de este trabajo, están relacionados con cada oferta de empleo revisada.

4.3.1. Clasificación con etiquetas determinadas

Como se comentó previamente, el objetivo de esta etapa es hacer que la herramienta escogida de Inteligencia Artificial, analice cada una de las ofertas de empleo para así relacionarlas con los resultados de aprendizaje determinados en el levantamiento inicial de este trabajo. Al ser 37 resultados de aprendizaje, se espera que cada oferta de empleo obtenga de 0 a 37 relaciones identificadas.

En esta etapa se utilizan 2 entradas. Por un lado, se ingresa al programa el archivo csv con las ofertas de trabajo filtradas y limpias. La segunda entrada utilizada es la lista de etiquetas que representan a cada resultado de aprendizaje, obtenida en la primera etapa de esta investigación.

Recordando la construcción de la base de datos, esta consta de cuatro columnas: Título, Empresa, Contenido y Perfil deseado. Para el objetivo de clasificación, se consideraran sólo tres de ellas, dejando fuera la columna Empresa.

Como las entradas para realizar una solicitud a la API de OpenAI son cadenas de texto, se implementa una función que concatena las tres columnas utilizadas en una sola cadena de texto. La función es aplicada al conjunto de datos completo. La cadena de texto resultante para cada oferta de empleo se determina de la siguiente forma:

```
1 descripciones = []
2 for index,row in df_ofertas.iterrows():
3     oferta = 'Titulo: ' + row['Titulo'] + '. Contenido oferta: ' + row['Contenido'] + '. Perfil
4     ↪ deseado: ' + row['Perfil']
5     descripciones.append(oferta)
```

Código 4.1: Concatenación de columnas del *dataset*.

Como se observa en el Código 4.1, cada una de las cadenas de texto generadas es guardada en una lista, la cual se añade a la base de datos original en una nueva columna llamada “descripciones”.

Por otro lado, como se mencionó previamente, se utilizan como etiquetas las determinadas en la primera sección de esta metodología, que representan a cada uno de los 37 resultados de aprendizaje utilizados.

- | | |
|--|---|
| 1. Análisis de Desafío Innovación | 7. Elaboración de Proyectos de Ingeniería |
| 2. Identificación de Problemas y Oportunidades | 8. Evaluación de Costos y Tiempos de Producción |
| 3. Diseño de Prototipos Innovadores | 9. Factibilidad Económica de Proyectos |
| 4. Formulación de Proyectos de Inversión | 10. Elaboración de Documentación Técnica |
| 5. Cálculo de Rentabilidad de Proyectos | 11. Metodologías Ágiles en Diseño de Soluciones |
| 6. Evaluación de Proyectos de Inversión | |

- | | |
|---|--|
| 12. Uso de Herramientas de Emprendimiento | 25. Orientación al Propósito Colectivo |
| 13. Evaluación de Oportunidades y Prototipos | 26. Comunicación Efectiva de Proyectos |
| 14. Análisis de Problemas Operativos | 27. Reevaluación de Propuesta de Valor |
| 15. Validación Comercial de Prototipos | 28. Testeo de Soluciones con Maquetas |
| 16. Estimación de Mercado y Barreras | 29. Evaluación Crítica de Acciones |
| 17. Construcción de Plan de Negocios | 30. Identificación y Solución de Problemas de Equipo |
| 18. Diseño de Modelos de Negocio | 31. Evaluación Económica y Social de Productos |
| 19. Prospección de Necesidades Sociales | 32. Validación de Producto Mínimo Viable |
| 20. Validación de Tracción de Proyecto | 33. Pitch y Exposiciones Orales |
| 21. Trabajo en Equipo en Proyectos Tecnológicos | 34. Exposición de Proyectos de Inversión |
| 22. Redacción de Informes de Proyecto | 35. Producción de Reportes Técnicos |
| 23. Presentación de Soluciones Técnicas | 36. Presentación con Recursos Audiovisuales |
| 24. Creación y Gestión de Equipos | 37. Evaluación Integral de Soluciones |

Esta lista de resultados de aprendizaje se guarda en una lista de Python.

Para llevar a cabo esta etapa, nuevamente es utilizada la API de GPT, modelo 4o. En esta ocasión las entradas de contexto e indicación a la API son más específicas que para la etapa de filtrado:

1. Cadena de contexto (*system*): *Eres un asistente experto en análisis de ofertas de trabajo y resultados de aprendizaje de un plan de formación universitario. Se te dará una oferta de trabajo y una lista de resultados de aprendizaje. Tu tarea es analizar la descripción de la oferta de empleo y determinar cuáles de los resultados de aprendizaje se encuentran involucrados en los requisitos de cada oferta.*
2. Indicación o *prompt (user)*: *Analiza la siguiente oferta de trabajo: oferta. Debes revisar uno a uno los 37 resultados de aprendizaje de la siguiente lista: {'', '.join(RA)} y determinar si el resultados de aprendizaje se ve involucrado en la descripción de la oferta o no.*

La configuración de tu respuesta debe ser solamente una lista de largo 37 del tipo [1,0,1,1,1,1,0,0,0,0,0,1,1,0,1,0,1,1,0,1,1,1,0,0,0,0,0,0,1,1,0,1], donde 1 indica que el resultado de aprendizaje se ve involucrado en la oferta, y 0 indica que no. Es importante que la lista siga el mismo orden que la lista de los resultados de aprendizaje.

No agregues información adicional, comentarios adicionales, ni introducción a tu respuesta. Responde únicamente la lista indicada.

Como se puede observar en las cadenas de texto ingresadas, en la entrada correspondiente al *prompt*, se incluye la lista de etiquetas a utilizar y el formato de salida objetivo. Se puede observar además, que se le pide al modelo no agregar información adicional, puesto que es necesario obtener una salida limpia con el estilo de una lista de Python, para poder procesar de mejor manera los resultados.

Así, la salida esperada es una cadena de texto con el formato visual de una lista de Python, compuesta de 37 elementos (1s o 0s).

Al procesar los resultados entregados por el modelos de GPT-4o para la tarea determinada previamente, se obtiene un tipo de anomalía (2 filas con una palabra adicional, o con símbolos extras). Para enfrentar esos casos se crea la función *clean_cell*, que elimina los valores que no son 0 o 1.

Finalmente, la matriz resultante es convertida al tipo de datos *integer*, asegurando sólo valores numéricos en el conjunto de datos.

4.3.2. Análisis preliminar de los resultados

Antes de comenzar con la preparación de los datos para la aplicación de los algoritmos de agrupación, se lleva a cabo una etapa de análisis de los primeros resultados obtenidos, con el objetivo de obtener los primeros resultados a nivel estadístico.

Considerando que cada relación positiva entre un resultado de aprendizaje y una oferta de empleo se marca con un valor igual a 1, para obtener la aparición total de cada etiqueta, se realiza la suma por columnas. Con esta información se obtiene la frecuencia absoluta de cada etiqueta.

Realizando la suma por filas se obtienen los siguientes datos estadísticos:

- Promedio de etiquetas por oferta.
- Desviación estándar de los resultados.
- Cantidad máxima de etiquetas relacionadas a una oferta.
- Cantidad mínima de etiquetas relacionadas a una oferta.

Estos resultados son acomodados en gráficos y tablas.

4.4. Construcción de la propuesta

Luego de haber obtenido los resultados de aprendizaje que responden a cada una de las ofertas de trabajo analizadas, el objetivos de esta etapa es lograr agrupar los resultados de aprendizajes en base a la data recolectada, con miras a generar la propuesta final. Esta etapa es abordada mediante 2 métodos. El primero, consiste en la aplicación de algoritmos de agrupación tradicionales, en particular se utiliza K-Means y agrupación jerárquica. El segundo enfoque utilizado es realizar el requerimiento a la API de GPT-4o. Así, esta etapa se divide en 3:

1. Preprocesamiento de los datos.
2. Aplicación de algoritmos de agrupación.
3. Agrupación mediante API GPT-4o.

4.4.1. Preprocesamiento de los datos

Para poder incorporar la información de la descripción de los resultados de aprendizaje a los métodos de agrupación, es necesario utilizar técnicas de procesamiento de lenguaje natural. En particular, se utilizan los *embeddings* para traducir el lenguaje natural a vectores numéricos que pueden ser procesados por algoritmos de agrupación utilizados.

Se utiliza el modelo especializado en *embeddings* de OpenAI. Esta decisión es motivada por la inclusión de forma nativa en el modelo de los procesos de “tokenización”, lematización y eliminación de *stopwords*, por lo que para utilizarlo no es necesario preprocesar los datos para el ingreso de los textos al modelo.

Las descripciones de los 37 resultados de aprendizaje son guardado en un *dataframe*, a partir del cual es seleccionada únicamente la columna correspondiente a las descripciones de los resultados de aprendizaje.

El Código 4.2 muestra la sección del código que realiza la llamada a la API y cómo la respuesta del modelo es procesada para obtener el vector de características deseado.

```
1 # Generar embeddings para cada texto
2 response = client.embeddings.create(
3     input=df2['Resultados de Aprendizaje'],
4     model="text-embedding-3-large")
5
6 # Extraer los embeddings de la respuesta
7 embeddings = []
8 for embedding in response.data:
9     emb = embedding.embedding
10    embeddings.append(emb)
11
12 df_embeddings = pd.DataFrame(embeddings)
13 df_embeddings
```

Código 4.2: Llamada al modelo “text-embeddings-large” de embeddings de OpenAI.

Luego de obtener la respuesta del modelo para cada resultado de aprendizaje, es extraído sólo el vector de *embeddings* para cada uno y, posteriormente, son guardado en un nuevo *dataframe*.

Como resultado de la sección anterior (etiquetado de la base de datos) se obtiene una matriz de unos y ceros, de 37 filas y 2.467 columnas, donde cada fila corresponde a un resultado de aprendizaje y cada columna representa una oferta de empleo analizada. Utilizando esta matriz, se genera un vector de pesos, el cual representa la demanda de cada resultado de aprendizaje en el estudio del mercado laboral. Para el cálculo de este vector, se ejecuta la

suma por filas.

Este vector de pesos es normalizado utilizando la función `StandardScaler()` de `sklearn`, obteniendo finalmente un vector de 37 filas y 1 columna.

El siguiente paso, es unificar los dos conjuntos de información disponible. Para esto, cada columna de la matriz de *embeddings* es escalada por el vector de pesos, logrando así traspasar la relevancia de cada resultado de aprendizaje en el mercado, a los datos que serán agrupados en la siguiente etapa.

4.4.2. Aplicación de algoritmos de agrupación

Se aplican dos algoritmos de agrupación:

1. K-Means
2. Agrupación jerárquica

Para los dos tipos clustering utilizados, se escalan los datos mediante *StandardScaler()*. Adicionalmente, se aplica PCA con el objetivo de reducir la dimensionalidad del conjunto de datos, y así poder visualizar los resultados de mejor manera, tal como se muestra en el Código 4.3.

```
1 # Escalar los datos
2 scaler = StandardScaler()
3 data_scaled = scaler.fit_transform(embeddings_ajustados)
4
5 # Aplicar PCA
6 pca = PCA(n_components=2)
7 principal_components = pca.fit_transform(data_scaled)
```

Código 4.3: Escalamiento de datos y PCA. Elaboración propia.

4.4.2.1. K-Means

El modelo utilizado para aplicar K-Means es el provisto por *sklearn*. En este trabajo, se determina utilizar tres parámetros de entrada para realizar el ajuste correspondiente:

- `n_clusters`: número de clusters a construir por el algoritmo. Determinado mediante el método del codo.
- `n_init`: número de veces que el algoritmo es iniciado con centroides diferentes. Si el número es mayor, aumenta la probabilidad de capturar la mejor combinación. Se configuró con 100 eventos de iniciación.
- `random_state`: asegura poder reproducir los resultados, manteniendo siempre los mismos componentes aleatorios del modelo. Se configuró igual a 42.

Para calcular la cantidad óptima de clústeres, se utiliza el método del codo, calculando la inercia de los resultados al entrenar el modelo con n cantidad de clústeres, siendo n un entero

entre 1 y 14, ambos inclusivos.

Finalmente, se vuelve a entrenar el modelo utilizando el n óptimo, determinado por el método del codo, y se generó la visualización de ellos. Además, se recuperan los datos de los clústeres correspondientes a cada resultado de aprendizaje, y se incluyen los centroides.

4.4.2.2. Agrupación Jerárquica

El segundo método de agrupación aplicado es la agrupación jerárquica.

Se utilizan las funciones *linkage()* para generar el proceso de agrupación y *dendrogram()* para configurar correctamente el árbol jerárquico. Los parámetros configurados son los siguientes:

- **method**: usado en *linkage()*. Indica el algoritmo usado para calcular las distancias entre clústers. Configurado con valor igual a *ward*.
- **Z**: parámetro de *dendrogram()*. Es la matriz de enlace calculada con *linkage()*. Contiene la información de las uniones del proceso de agrupación.

Finalmente, con la función *fcluster()*, se configura el criterio para realizar el corte de clústeres en la jerarquía creada. En este caso el criterio es el número de clústeres igual a 4.

4.4.3. Agrupación mediante API GPT-4o

Además de los procesos de clusterización revisados previamente, se lleva a cabo una última forma de generar agrupaciones de resultados de aprendizaje, con el objetivo de entregar información relevante para la futura construcción de los cursos enfocados a innovación en el plan de formación de Ingeniería Civil Eléctrica.

Se utiliza la API de OpenAI, en particular se llama al modelo GPT-4o. Como fue revisado previamente, la llamada se debe realizar creando dos *prompt* principales, uno de contexto y uno con la indicación de la tarea a realizar por el modelo. Además, el objetivo de utilizar la API es poder ingresar al modelo toda la información recolectada a lo largo de este estudio para que genere los resultados solicitados.

El *dataframe* final, utilizado para este objetivo contiene las siguientes columnas: Resultados de Aprendizaje, Etiquetas, Competencias generales, Competencia, Subcompetencia y Peso Escalado. Todas estas columnas han sido descritas previamente.

El contexto que se ingresa al modelo es el siguiente: *Eres un asistente experto en organización de planes de formación universitaria en innovación. Se te entregará una tabla con varios parámetros que describen distintos resultados de aprendizaje, añadiendo una investigación sobre el mercado laboral. Debes generar una clusterización para organizar los cursos enfocados en innovación de la carrera Ingeniería civil Eléctrica.*

Por otro lado, el *prompt* con la indicación de la tarea específica a realizar por el modelo, es el mostrado en el Código 26, donde además, se muestra la función completa utilizada para la llamada a la API.

```

1 def clusterizacion_gpt(df_completo,system_context):
2     prompt = f"""Con la siguiente tabla: {df_completo} debes organizar una propuesta de
3         ↳ plan de formación, agrupando los resultados de aprendizaje. Las descripciones de las
4         ↳ columnas de la tabla son:
5
6     Resultado de Aprendizaje: descripción del resultado de aprendizaje.
7     Etiqueta: nombre para identificar cada resultado de aprendizaje.
8     Competencias generales: grupos de competencias orientadas al saber (Conocimiento),
9         ↳ saber hacer (Habilidades) y saber ser (Actitudes).
10    Competencias: Competencias que agrupan varias subcompetencias del plan de formación.
11    Subcompetencias: Subcompetencias que agrupan a los resultados de aprendizaje finales.
12    Peso Escalado: Información obtenida del análisis de ofertas de mercado laboral. Simboliza
13         ↳ la demanda de cada resultado de aprendizaje en el mercado laboral. Se encuentra
14         ↳ escalado con StandarScaler.
15
16    El objetivo es agrupar los resultados de aprendizaje considerando toda la información de
17         ↳ la tabla, incluyendo los pesos del mercado laboral, para generar propuestas de cursos
18         ↳ enfocados en
19    innovación que se puedan incluir en el plan curricular de la carrera. Para esto el resultado
20         ↳ debe organizarse en 3 a 5 cursos totales.
21    No agregues un curso especializado en comunicación y trabajo en equipo, estos resultados
22         ↳ de aprendizaje se deben incluir en todos los otros cursos para ser desarrollados de
23         ↳ forma progresiva.
24    Si lo consideras necesario, algunos resultados de aprendizaje se pueden repetir.
25    Es importnte usar todos los resultados de aprendizaje de la lista (37) pero si consideras
26         ↳ necesario eliminar alguno, indícalo.
27    Identifica el índice del resultado de aprendizaje para asegurar que estén todos."""
28
29    response = client.chat.completions.create(
30        model='gpt-4o',
31        messages=[
32            {"role": "system", "content": system_context},
33            {"role": "user", "content": prompt}
34        ]
35    )
36    return response.choices[0].message.content.strip()

```

Código 4.4: Función para llamada a la API de OpenAI para agrupar resultados de aprendizaje.

Como se observa en el Código 4.4, la entrada considera el *dataframe* completo.

Los resultados de esta agrupación son guardados en un documento, a partir del cual son construidas distintas visualizaciones con la respuesta entregada.

4.5. Consolidación de la propuesta

A partir de los resultados obtenidos, se realiza el análisis final de los 3 tipos de agrupación. Tras el análisis detallado, se genera una propuesta, considerando la aplicabilidad al programa actual.

Capítulo 5

Resultados

Este capítulo presenta los resultados obtenido en cada una de las etapas descritas previamente en el Capítulo 4. Se acompañan del análisis correspondiente a cada uno.

5.1. Recolección de datos

Dado que la recolección de datos para este trabajo de título se abordó desde dos frentes, la presentación de los resultados se dividirá también en dos subsecciones: Análisis del currículum de Ingeniería Civil Eléctrica y Recolección de datos web.

5.1.1. Análisis del currículum de Ingeniería Civil Eléctrica

La Tabla 5.1 muestra las competencias, tanto específicas (CE) como genéricas (CG) que se ven abordadas en cada uno de los cuatro cursos obligatorios relacionados a la formación en innovación de la carrera de Ingeniería Civil Eléctrica. Se marcan con 1 las competencias abordadas y, con 0, las no abordadas en el curso. Además, las columnas de color gris son las correspondientes a las competencias que se consideraron relevantes para la formación en innovación.

Tabla 5.1: Resultados análisis currículum universitario

Código	CE1	CE2	CE3	CE4	CE5	CE6	CE7	CG1	CG2	CG3	CG4	CG5	CG6	CG7
EL3105	0	0	0	0	0	1	1	1	0	0	0	0	1	0
EL4204	0	0	0	1	0	0	1	1	0	1	0	0	0	0
EL6101	0	0	0	1	1	1	1	1	0	0	1	1	1	1
EL6201	0	1	0	0	1	1	0	1	0	0	1	0	1	1

Tras haber analizado esta distribución de competencias a las que tributan los cursos obligatorios de Ingeniería Eléctrica que abarcan tópicos de innovación, se obtiene un total de 37 resultados de aprendizajes diferentes.

Estos resultados de aprendizaje, se muestran en la Tabla A.1, relacionados a un índice, que será el correspondiente a cada resultado de aprendizaje para todo este trabajo de título. Además, se añade la columna de etiquetas generadas por GPT 4o para representar cada uno de estos resultados de aprendizaje.

A partir de la Tabla 5.1, y el mapeo de resultados de aprendizaje, se generan las Figuras

5.1 y 5.2, que muestran redes multipartitas con las relaciones entre las competencias (específicas y genéricas, respectivamente), los cuatro cursos obligatorios analizados y los resultados de aprendizaje identificados.

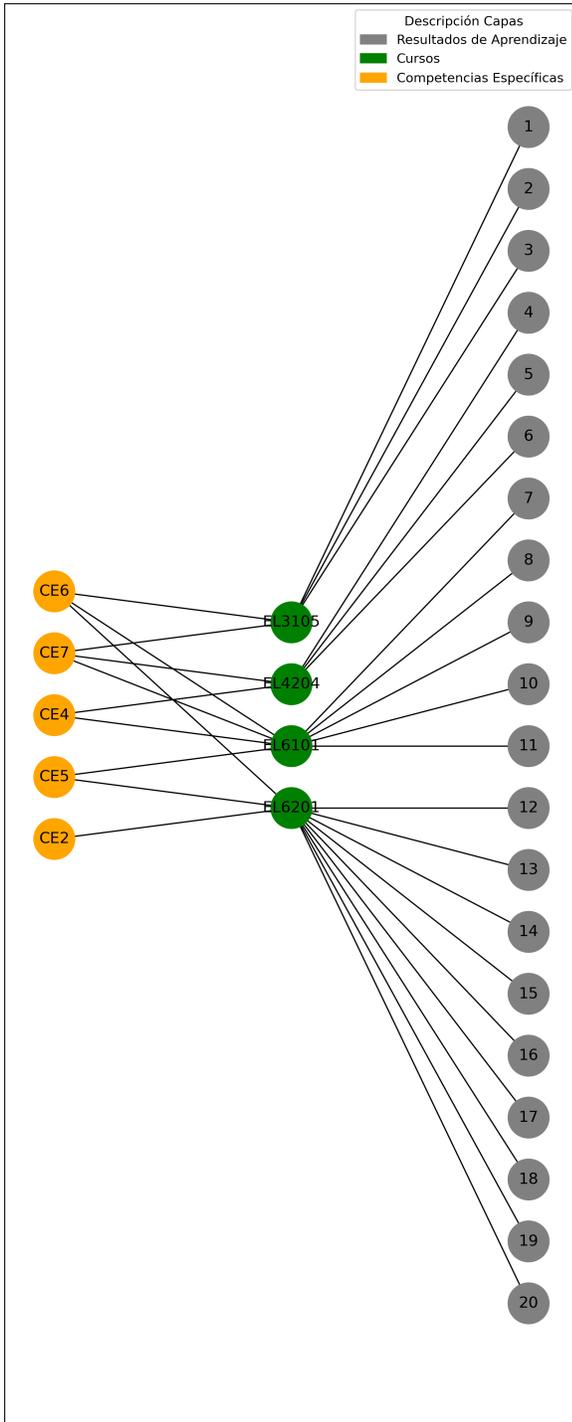


Figura 5.1: Relación Competencias Específicas - Curso - Resultado de Aprendizaje.

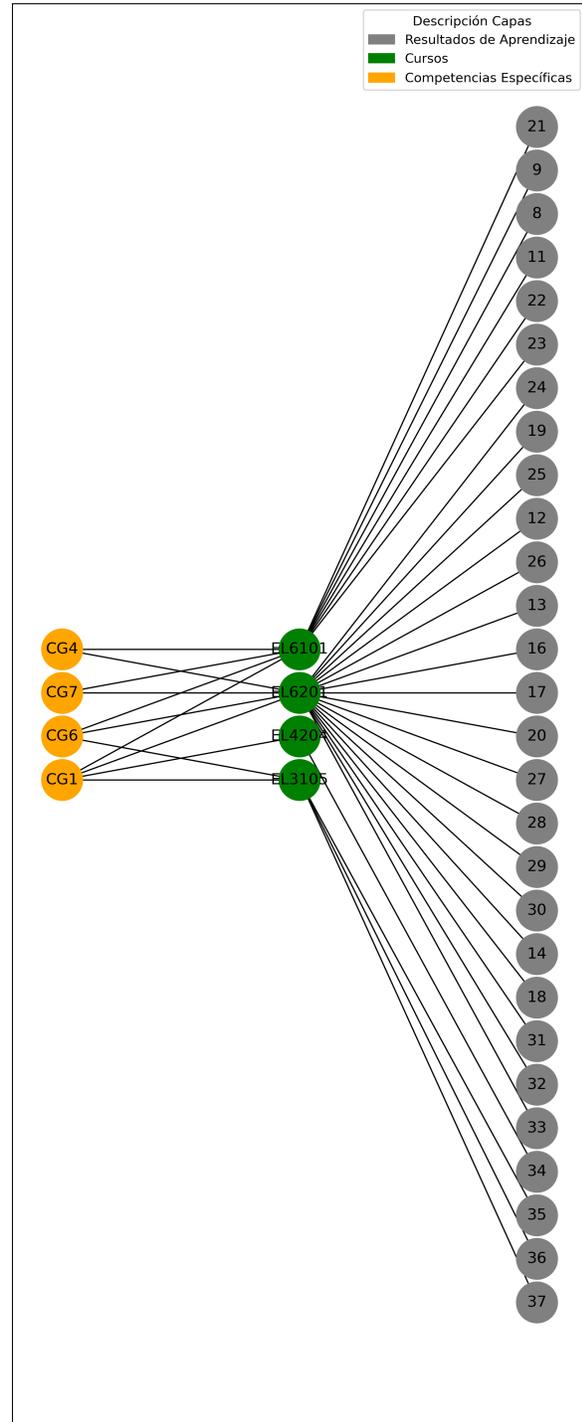


Figura 5.2: Relación Competencias Genéricas - Curso - Resultado de Aprendizaje.

La Tabla 5.2 contiene los resultados generados por GPT-4o para la solicitud de organizar los resultados de aprendizaje en competencias y subcompetencias. Se añadió además, en la primera columna, el grupo de competencia asociado al resultado de aprendizaje, según GPT-4o.

Tabla 5.2: Organización en los distintos niveles: Grupo de competencia, Competencia, Subcompetencia y Resultado de Aprendizaje.

Grupo de Competencia	Competencia	Subcompetencia	Resultado de Aprendizaje	N
Conocimiento	Diseño y prototipado de soluciones	Diseño de Prototipos	Formulación de Proyectos de Inversión	4
Conocimiento	Evaluación de Prototipos y Productos	Evaluación de Prototipos y Productos	Cálculo de Rentabilidad de Proyectos	5
Conocimiento	Diseño y prototipado de soluciones	Evaluación de Prototipos y Productos	Evaluación de Proyectos de Inversión	6
Conocimiento	Gestión de Proyectos y Emprendimiento	Planificación y Evaluación de Proyectos	Evaluación de Costos y Tiempos de Producción	8
Conocimiento	Gestión de Proyectos y Emprendimiento	Planificación y Evaluación de Proyectos	Factibilidad Económica de Proyectos	9
Conocimiento	Comunicación y Documentación	Documentación Técnica	Elaboración de Documentación Técnica	10
Conocimiento	Gestión de Proyectos y Emprendimiento	Herramientas y Métodos de Emprendimiento	Metodologías Ágiles en Diseño de Soluciones	11
Conocimiento	Gestión de Proyectos y Emprendimiento	Herramientas y Métodos de Emprendimiento	Uso de Herramientas de Emprendimiento	12
Conocimiento	Gestión de Proyectos y Emprendimiento	Herramientas y Métodos de Emprendimiento	Estimación de Mercado y Barreras	16
Conocimiento	Identificación y análisis de problemas	Análisis Contextual y del Desafío	Prospección de Necesidades Sociales	19
Conocimiento	Identificación y análisis de problemas	Evaluación y Validación de Problemas	Validación de Tracción de Proyecto	20
Conocimiento	Diseño y prototipado de soluciones	Evaluación de Prototipos y Productos	Evaluación Económica y Social de Productos	31
Conocimiento	Diseño y prototipado de soluciones	Evaluación de Prototipos y Productos	Evaluación Integral de Soluciones	37
Habilidades	Identificación y análisis de problemas	Análisis Contextual y del Desafío	Análisis de Desafío Innovación	1

(Continúa en la página siguiente)

Tabla 5.2: Continuación. Organización en los distintos niveles: Grupo de competencia, Competencia, Subcompetencia y Resultado de Aprendizaje.

Grupo de Competencia	Competencia	Subcompetencia	Resultado de Aprendizaje	N
Habilidades	Identificación y análisis de problemas	Análisis Contextual y del Desafío	Identificación de Problemas y Oportunidades	2
Habilidades	Diseño y prototipado de soluciones	Diseño de Prototipos	Diseño de Prototipos Innovadores	3
Habilidades	Diseño y prototipado de soluciones	Diseño de Prototipos	Elaboración de Proyectos de Ingeniería	7
Habilidades	Gestión de Proyectos y Emprendimiento	Planificación y Evaluación de Proyectos	Evaluación de Oportunidades y Prototipos	13
Habilidades	Gestión de Proyectos y Emprendimiento	Planificación y Evaluación de Proyectos	Análisis de Problemas Operativos	14
Habilidades	Diseño y prototipado de soluciones	Evaluación de Prototipos y Productos	Validación Comercial de Prototipos	15
Habilidades	Gestión de Proyectos y Emprendimiento	Planificación y Evaluación de Proyectos	Construcción de Plan de Negocios	17
Habilidades	Gestión de Proyectos y Emprendimiento	Herramientas y Métodos de Emprendimiento	Diseño de Modelos de Negocio	18
Habilidades	Comunicación y Documentación	Presentación y Exposición	Presentación de Soluciones Técnicas	23
Habilidades	Diseño y prototipado de soluciones	Diseño de Prototipos	Testeo de Soluciones con Maquetas	28
Habilidades	Diseño y prototipado de soluciones	Evaluación de Prototipos y Productos	Validación de Producto Mínimo Viable	32
Habilidades	Comunicación y Documentación	Presentación y Exposición	Pitch y Exposiciones Orales	33
Habilidades	Comunicación y Documentación	Presentación y Exposición	Exposición de Proyectos de Inversión	34
Habilidades	Comunicación y Documentación	Documentación Técnica	Producción de Reportes Técnicos	35
Habilidades	Comunicación y Documentación	Presentación y Exposición	Presentación con Recursos Audiovisuales	36
Actitud	Trabajo en Equipo y Liderazgo	Colaboración y Gestión de Equipos	Trabajo en Equipo en Proyectos Tecnológicos	21
Actitud	Comunicación y Documentación	Documentación Técnica	Redacción de Informes de Proyecto	22
Actitud	Trabajo en Equipo y Liderazgo	Colaboración y Gestión de Equipos	Creación y Gestión de Equipos	24

(Continúa en la página siguiente)

Tabla 5.2: Continuación. Organización en los distintos niveles: Grupo de competencia, Competencia, Subcompetencia y Resultado de Aprendizaje.

Grupo de Competencia	Competencia	Subcompetencia	Resultado de Aprendizaje	N
Actitud	Trabajo en Equipo y Liderazgo	Colaboración y Gestión de Equipos	Orientación al Propósito Colectivo	25
Actitud	Comunicación y Documentación	Presentación y Exposición	Comunicación Efectiva de Proyectos	26
Actitud	Identificación y análisis de problemas	Evaluación y Validación de Problemas	Reevaluación de Propuesta de Valor	27
Actitud	Trabajo en Equipo y Liderazgo	Evaluación y Mejora Continua	Evaluación Crítica de Acciones	29
Actitud	Trabajo en Equipo y Liderazgo	Evaluación y Mejora Continua	Identificación y Solución de Problemas de Equipo	30

(Fin de la tabla)

Finalmente, a partir de la Tabla 5.2, se generaron tres distintas redes multipartitas, una por cada grupo de competencias. La Figura 5.3, corresponde a los resultados de aprendizaje clasificados bajo el grupo de Conocimiento (saber saber). La Figura 5.4, corresponde a los resultados de aprendizaje clasificados bajo el grupo de Habilidades (saber hacer) y, finalmente, la Figura 5.5, corresponde a los resultados de aprendizaje clasificados bajo el grupo Actitud (saber ser).

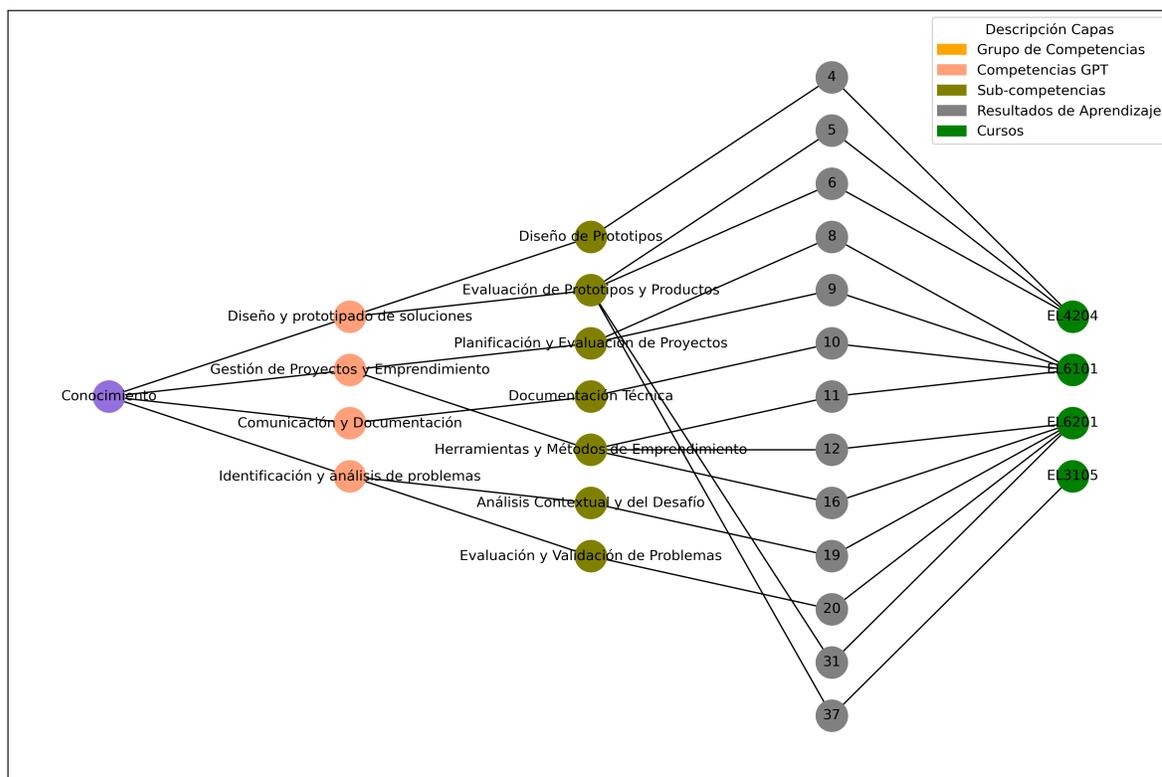


Figura 5.3: Red de conexiones entre competencias, subcompetencias, resultados de aprendizaje y cursos bajo el grupo de competencias relacionadas al conocimiento.

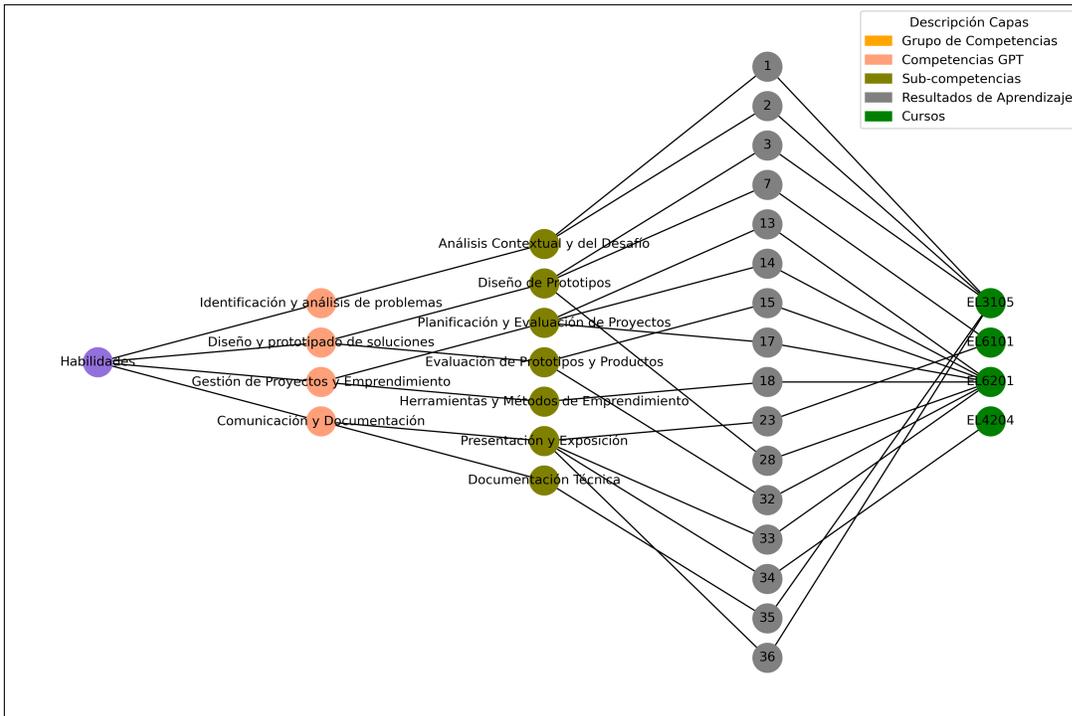


Figura 5.4: Red de conexiones entre competencias, subcompetencias, resultados de aprendizaje y cursos bajo el grupo de competencias relacionadas a las habilidades.

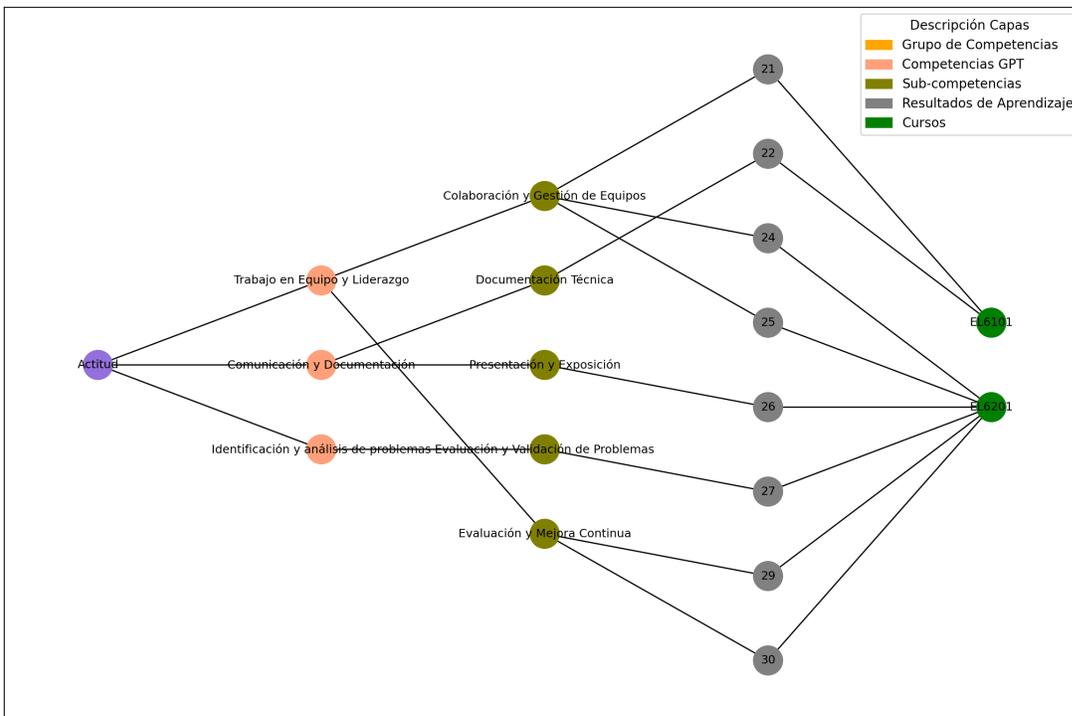


Figura 5.5: Red de conexiones entre competencias, subcompetencias, resultados de aprendizaje y cursos bajo el grupo de competencias relacionadas a la actitud.

5.1.2. Recolección de ofertas de trabajo web

A partir del trabajo realizado para recuperar ofertas de empleo alojadas en la web de *trabajando.com*, se obtuvieron los resultados de la Tabla 5.3. En ella se muestra la cantidad de ofertas recuperadas por cada una de las búsquedas, alcanzando un total de 4.499 ofertas de empleo recuperadas.

Tabla 5.3: Cantidad de ofertas de trabajo recuperadas por búsqueda realizada.

Búsqueda	Cantidad Resultados
Ingeniero Eléctrico	734
Ingeniero Electrónico	593
Ingeniero Telecomunicaciones	192
Ingeniero de Proyectos Eléctricos	374
Analista de datos	300
Ingeniero de datos	911
Ingeniero TI	1.365
Inteligencia Artificial	30
Total	4.499

5.2. Preparación de los datos

La etapa de preparación de los datos para ser procesados con algoritmos de agrupación se lleva a cabo en tres partes:

1. Eliminación de duplicados y limpieza de datos.
2. Experimento para identificar el mejor algoritmo de clasificación, comparando *Zero-Shot* con los resultados de GPT 4o.
3. Eliminación final de registros identificados como no relevantes.

5.2.1. Comparación de modelos de clasificación

Una vez ejecutada la limpieza de datos y eliminación de registros duplicados, se obtuvo una base de datos compuesta por 4.487 observaciones.

Tras haber ejecutado ambos modelos con el set de datos correspondiente a la búsqueda “Inteligencia Artificial”, con el objetivo de realizar la clasificación entre la misma búsqueda y “Otros”, se obtuvieron las matrices de confusión de la Figura 5.6. En el lado izquierdo se muestra la matriz de confusión correspondiente al desempeño de GPT-4o y, a la derecha, la matriz correspondiente a la clasificación de *Bart-Large-mnli (Zero-Shot)*.

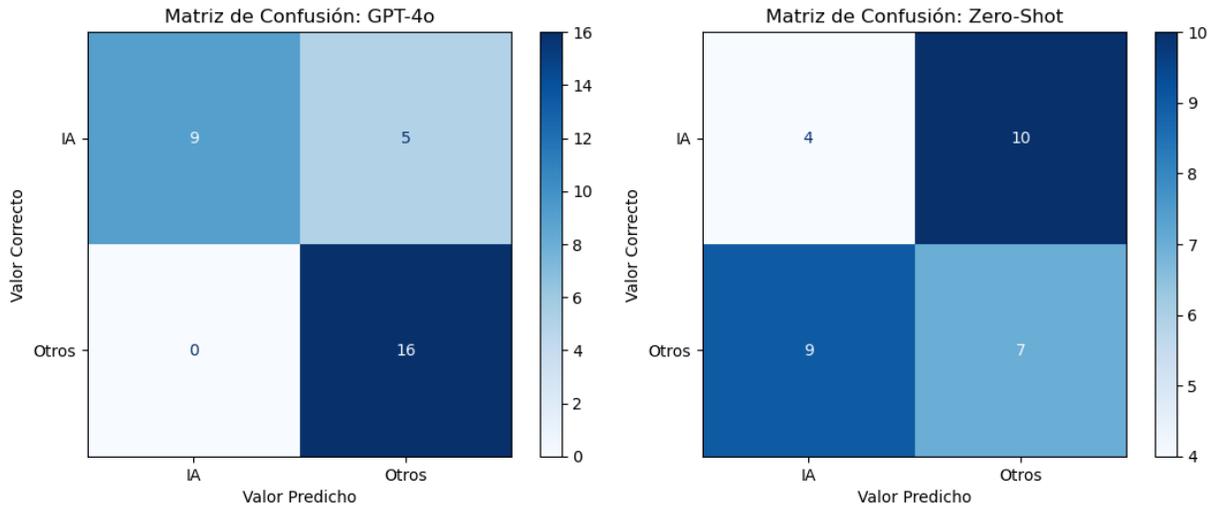


Figura 5.6: Matrices de confusión resultantes de la clasificación con GPT 4o (izquierda) y *Zero-Shot* (derecha).

La Tabla 5.4, ofrece un resumen de la cantidad de observaciones por clases mediante los tres métodos utilizados: manual, clasificación mediante *prompt* con GPT-4o y clasificación *Zero-Shot* con *bart-large-mnli*.

Tabla 5.4: Comparación de métricas de rendimiento de clasificación de textos con GPT-4o y *Zero-Shot*.

Etiqueta	Manual	GPT 4o	Zero-Shot
Inteligencia Artificial	14	9	13
Otros	16	21	17

La Tabla 5.5, muestra la comparación de las métricas utilizadas para evaluar el rendimiento de ambos modelos.

Tabla 5.5: Comparación de métricas de rendimiento de clasificación de textos con GPT-4o y *Zero-Shot*.

Métrica	GPT 4o	Zero-Shot
Accuracy	0,833	0,367
Precision	1	0,308
Recall	0,643	0,286
F1 - score	0.7826	0,296

5.2.2. Eliminación de registros no relevantes

En la Tabla 5.6, se muestran las características extraídas desde los paneles disponibilizados por OpenAI para controlar el uso de la API, incluyendo el costo y tiempo de ejecución.

Tabla 5.6: Resultados del rendimiento del programa.

Característica	Valor
Modelo	GPT-4o
Tiempo de ejecución	73 min 14,0 s
Costo	17,12 USD
Tokens de contexto	2.938.811
Tokens generados	161.911
Tokens totales	3.100.722

Luego de haber ejecutado la clasificación con GPT 4o, se obtienen los resultados de la Tabla 5.7, en la que se muestra la cantidad de observaciones clasificadas por cada búsqueda. Además, se comparan con la cantidad inicial obtenida en cada búsqueda. La Figura 5.7, muestra el gráfico de barras correspondiente a la Tabla 5.7, permitiendo visualizar la distribución de la clasificación.

Tabla 5.7: Comparación cantidad de ofertas clasificadas por GPT 4o con el estado inicial de ofertas de empleo.

Búsqueda	Cantidad inicial	Cantidad final
Ingeniero Eléctrico	734	722
Ingeniero Electrónico	593	71
Ingeniero Telecomunicaciones	192	141
Ingeniero de Proyectos Eléctricos	374	297
Analista de datos	300	521
Ingeniero de datos	911	55
Ingeniero TI	1.365	630
Inteligencia Artificial	30	31
Otros	-	2.019
Total	4.499	4.487

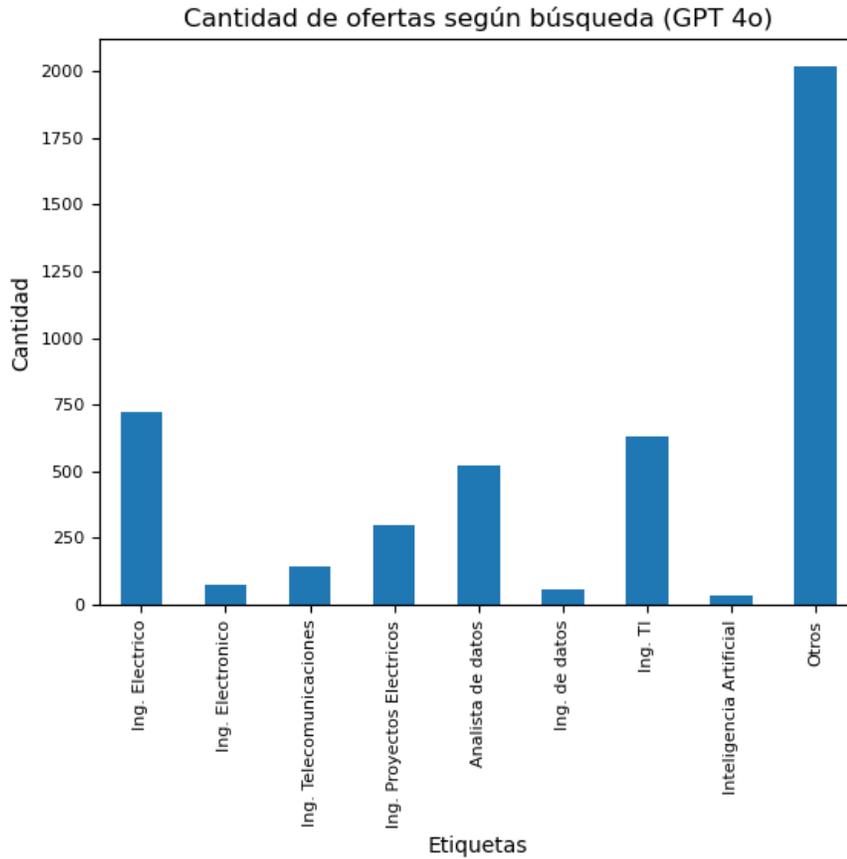


Figura 5.7: Visualización del resultado generado por GPT 4o en la clasificación de ofertas relevantes para estudio, mostrado en la Tabla 5.3.

5.3. Etiquetado de la base de datos

En la etapa de etiquetado de las ofertas de empleo de la base de datos según los resultados de aprendizaje utilizados en esta investigación, se presenta la Tabla 5.8, la cual contiene las características extraídas desde los paneles de la API de OpenAI.

Tabla 5.8: Resultados del rendimiento del programa al identificar resultados de aprendizaje en ofertas de trabajo.

Característica	Valor
Modelo	GPT-4o
Tiempo de ejecución	53 min 37,8 s
Costo	11,69 USD
Tokens de contexto	1.950.034
Tokens generados	126.150
Tokens totales	2.076.184

5.3.1. Análisis preliminar de los resultados

Antes de ingresar los resultados obtenidos del proceso de etiquetado a la etapa de agrupación, se realizó el análisis preliminar de los resultados. En primer lugar se obtuvo la Tabla

5.9, que muestra la cantidad de ofertas de empleo en las que fue identificado cada uno de los resultados de aprendizaje.

Tabla 5.9: Cantidad de ofertas de empleo en las que el modelo identifica cada resultado de aprendizaje.

N	Etiqueta	Cantidad
1	Análisis de Desafío Innovación	392
2	Identificación de Problemas y Oportunidades	2.337
3	Diseño de Prototipos Innovadores	30
4	Formulación de Proyectos de Inversión	255
5	Cálculo de Rentabilidad de Proyectos	197
6	Evaluación de Proyectos de Inversión	285
7	Elaboración de Proyectos de Ingeniería	1.303
8	Evaluación de Costos y Tiempos de Producción	550
9	Factibilidad Económica de Proyectos	104
10	Elaboración de Documentación Técnica	2047
11	Metodologías Ágiles en Diseño de Soluciones	304
12	Uso de Herramientas de Emprendimiento	14
13	Evaluación de Oportunidades y Prototipos	560
14	Análisis de Problemas Operativos	2.365
15	Validación Comercial de Prototipos	48
16	Estimación de Mercado y Barreras	92
17	Construcción de Plan de Negocios	169
18	Diseño de Modelos de Negocio	58
19	Prospección de Necesidades Sociales	222
20	Validación de Tracción de Proyecto	210
21	Trabajo en Equipo en Proyectos Tecnológicos	2.413
22	Redacción de Informes de Proyecto	2.023
23	Presentación de Soluciones Técnicas	1.478
24	Creación y Gestión de Equipos	1.008
25	Orientación al Propósito Colectivo	1.001
26	Comunicación Efectiva de Proyectos	1.896
27	Reevaluación de Propuesta de Valor	200
28	Testeo de Soluciones con Maquetas	43
29	Evaluación Crítica de Acciones	1.659
30	Identificación y Solución de Problemas de Equipo	1.483
31	Evaluación Económica y Social de Productos	288
32	Validación de Producto Mínimo Viable	100
33	Pitch y Exposiciones Orales	1.757
34	Exposición de Proyectos de Inversión	1.125
35	Producción de Reportes Técnicos	2.100
36	Presentación con Recursos Audiovisuales	203
37	Evaluación Integral de Soluciones	1.936

La Figura 5.8 presenta la visualización gráfica de la Tabla 5.9, permitiendo dimensionar visualmente las diferencias entre los resultados de aprendizaje.

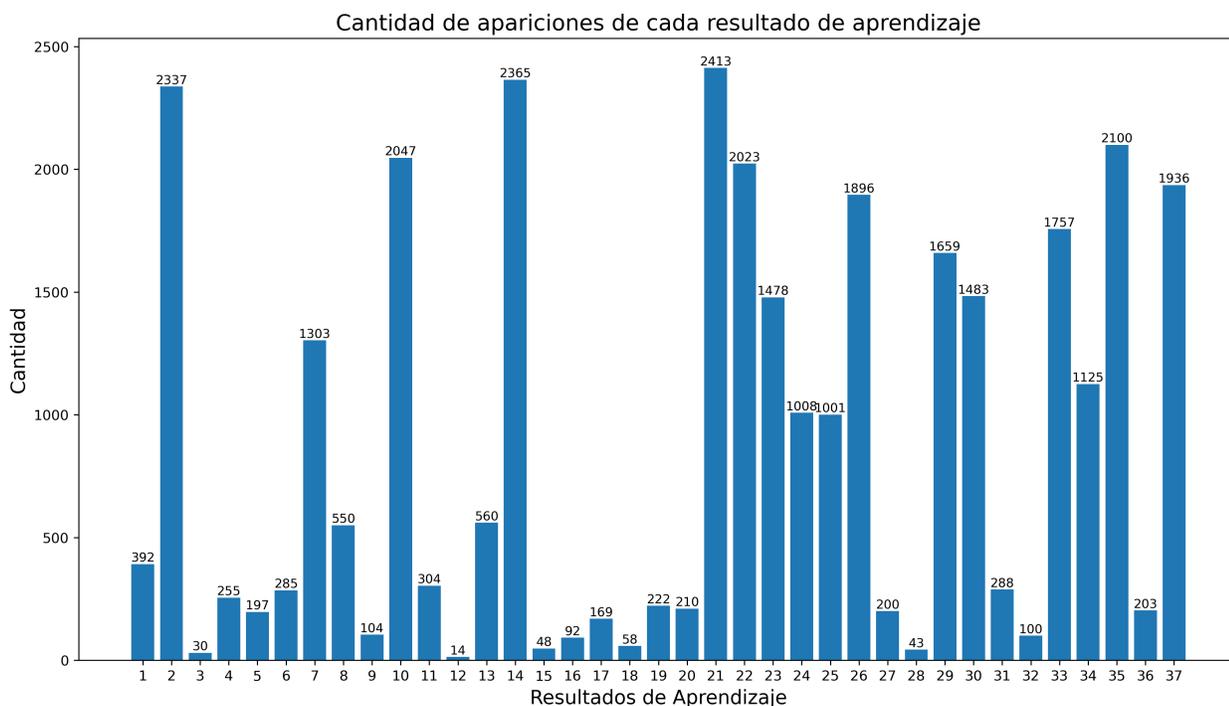


Figura 5.8: Visualización de la cantidad de veces que una resultado de aprendizaje fue identificado en alguna oferta de trabajo.

En la Tabla 5.10, se presentan las medidas estadísticas básicas necesarias para analizar los resultados, como el promedio, desviación estándar, cantidades máxima y mínima de resultados de aprendizaje identificados en una sola oferta y la cantidad de ofertas en las que se obtienen el máximo y el mínimo.

Tabla 5.10: Análisis general del resultado del etiquetado.

Medida	Valor
Promedio	13,0693
Desv. estándar	3,7778
Cantidad máxima de etiquetas	31
Nº de ofertas con el máximo	1
Cantidad mínima de etiquetas	1
Nº de ofertas con el mínimo	4

Finalmente, en la Figura 5.9, se observa en el eje horizontal la cantidad de resultados de aprendizaje identificados en una misma oferta de empleo y, en el eje vertical, la cantidad de ofertas en las que se identifica esa cantidad de resultados de aprendizaje. Por ejemplo, en 90 ofertas de empleo fueron identificados 8 resultados de aprendizaje.

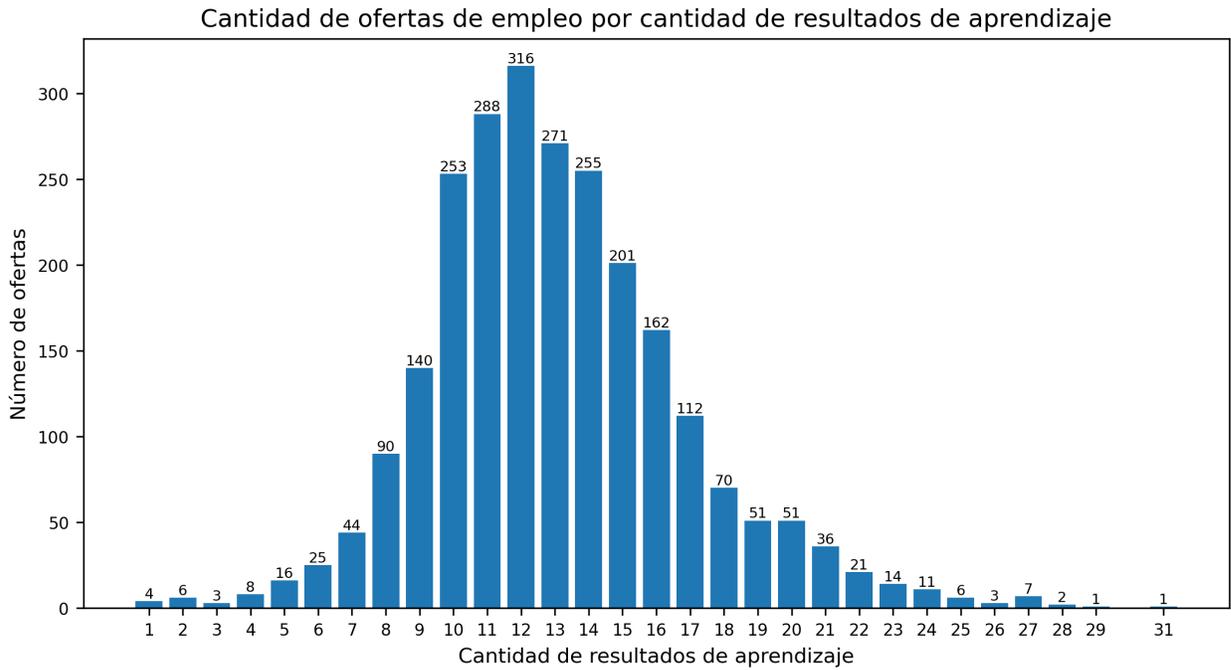


Figura 5.9: Visualización de la cantidad de ofertas que tienen la misma cantidad de resultados de aprendizaje identificados.

5.4. Construcción de la propuesta

Luego de haber preprocesado los datos mediante *embeddings*, se procedió con el escalamiento y reducción de dimensionalidad mediante PCA. Realizados estos pasos, se aplicaron dos algoritmos de agrupación.

El primer algoritmo fue K-Means. Para determinar el número de clúster óptimo para ser ingresado al algoritmo, se utilizó el método del codo, obteniendo el resultado de la Figura 5.10 ($n_clusters = 4$).

Se reentrenó el modelo con el parámetro encontrado anteriormente, obteniendo el resultado de la Figura 5.11. En esta, se muestran los clústeres formados identificados con colores. Adicionalmente, las equis de color rojo muestran los centroides definidos por el algoritmo para cada uno de los clústeres.

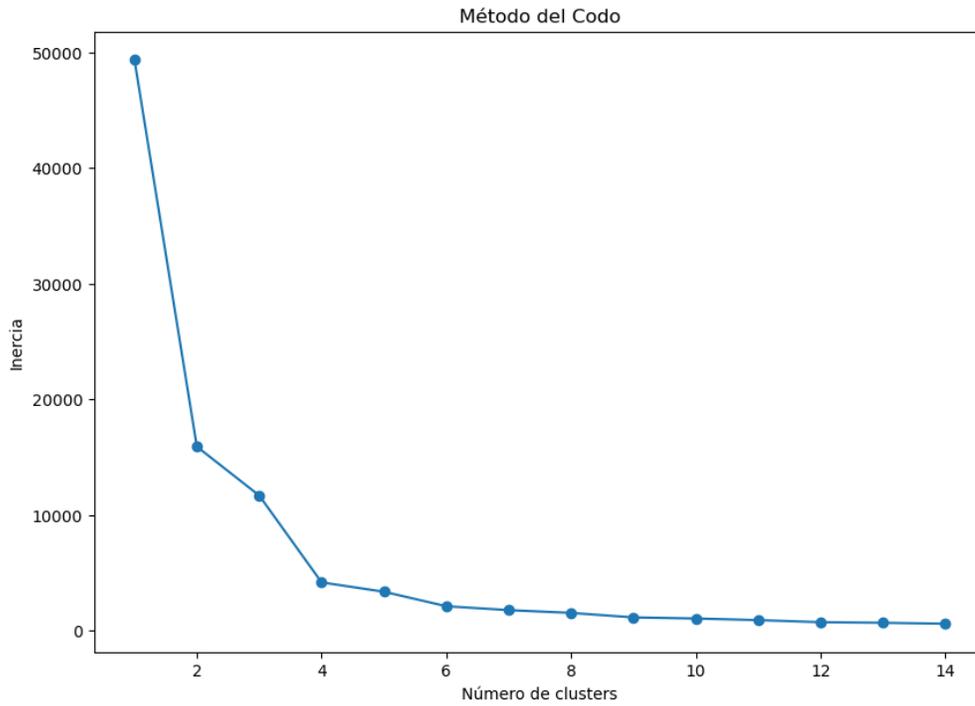


Figura 5.10: Método del codo aplicado para determinar clústeres óptimos en K-Means.

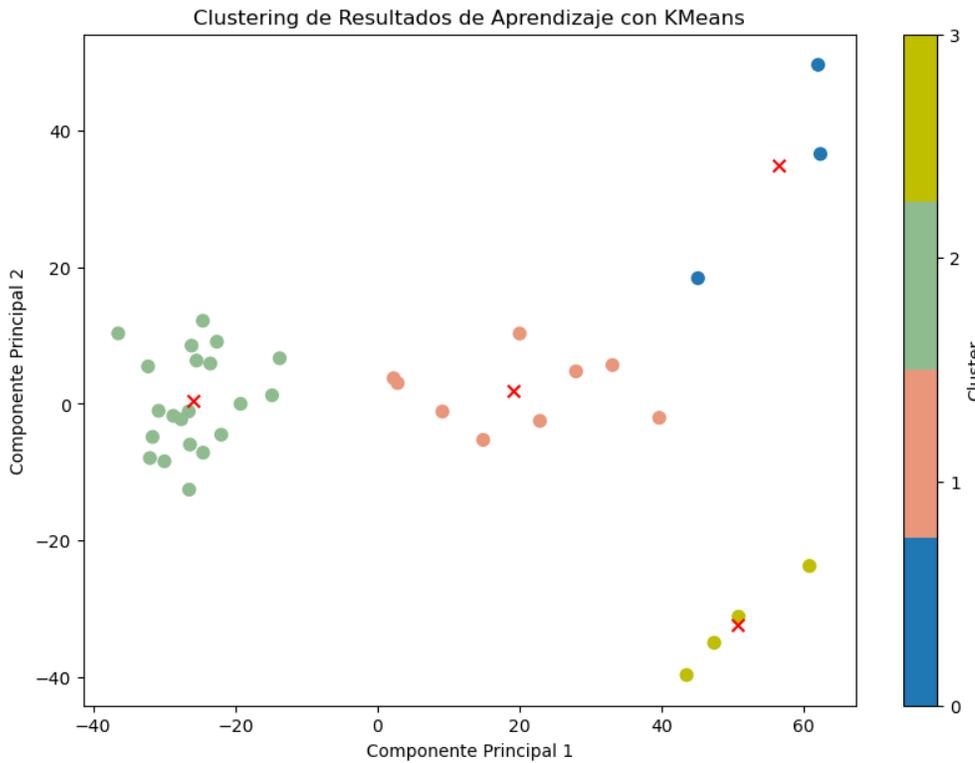


Figura 5.11: Resultados de agrupación con K-Means.

El segundo algoritmo utilizado fue la agrupación jerárquica de estrategia aglomerativa. La

Figura 5.12, muestra el resultado obtenido. La línea horizontal marca el corte en la distancia determinada (distancia = 40) para la formación de 4 clústeres, indicado cada agrupación de un color distinto. Con esto, se observa que este algoritmo genera 6 clústeres diferentes.

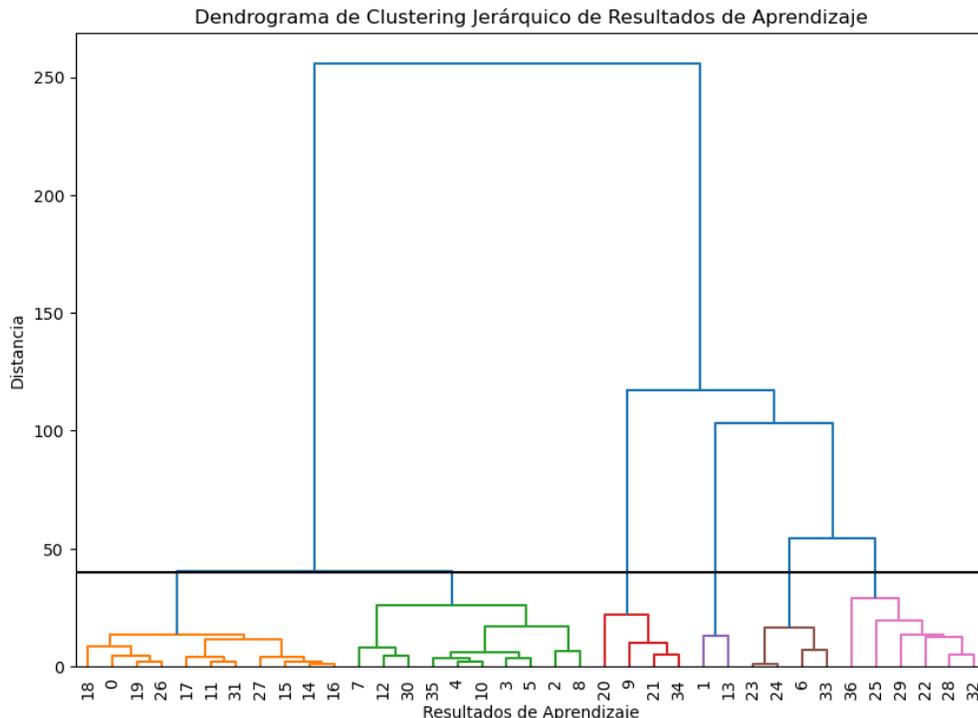


Figura 5.12: Resultados de agrupación jerárquica aglomerativa.

Al ingresar los datos recolectados a la API de GPT-4o, la respuesta fue bastante completa. Sin pedir estas especificaciones, el modelo respondió de forma adicional entregando un nombre y el enfoque de cada curso construido. La respuesta organizó los resultados de aprendizaje en 5 cursos diferentes. En la Tabla 5.11 se proporciona el título del curso entregado por GPT-4o y la descripción del enfoque de cada curso.

Tabla 5.11: Nombre y enfoque de cursos entregados por GPT-4o.

N	Nombre del curso	Enfoque
1	Análisis e Identificación de Problemas y Oportunidades Innovadoras	En este curso se trabaja en la identificación y análisis de problemas y oportunidades, con énfasis en entender y contextualizar los desafíos. También incluye competencias de comunicación y trabajo en equipo integradas en los proyectos.
2	Diseño y Prototipado de Soluciones Innovadoras	Este curso se enfoca en diseñar y desarrollar prototipos y evaluar sus funcionalidades utilizando metodologías ágiles.
3	Evaluación y Gestión de Proyectos Innovadores	Aquí se abordan las herramientas y metodologías para la gestión y evaluación de proyectos con criterios de eficiencia y sostenibilidad.
4	Implementación y Validación de Proyectos	En este curso se enfoca en la implementación, monitoreo y validación de proyectos, evaluando sus impactos y realizando iteraciones.
5	Monitoreo y Evaluación Continua de Proyectos	Enfocado en la evaluación continua y monitoreo para asegurar el éxito del proyecto hasta su fase de consolidación y posterior mejora.

La respuesta exacta de GPT-4o se encuentra en el Anexo A.2.1 de este trabajo de título.

Finalmente, se presenta la Tabla 5.12, que muestra cada resultado de aprendizaje, asociado al clúster que le fue asignado en cada uno de los procesos de agrupación anterior. La Figura 5.13 contiene una demostración visual de las agrupaciones creadas por los tres diferentes métodos de agrupación utilizados (K-Means, Jerárquico y GPT-4o según el orden de la figura). Se destaca la fila 32 de la tabla, que en la columna GPT-4o se indica cluster 5 con un asterisco. Esta marca señala una particularidad: el modelo no incluyó directamente este resultado de aprendizaje, sino que añadió un nuevo resultado a la lista y lo agregó al grupo 5. Al revisarlo, este nuevo resultado de aprendizaje es similar al número 32, por lo que se consideró en su lugar.

Tabla 5.12: Clúster asignado a cada resultado de aprendizaje para los tres tipos de agrupación utilizada.

N	Etiqueta	K-Means	Jerárquico	GPT-4o
1	Análisis de Desafío Innovación	1	1	1
2	Identificación de Problemas y Oportunidades	0	4	1
3	Diseño de Prototipos Innovadores	1	2	2
4	Formulación de Proyectos de Inversión	1	2	2
5	Cálculo de Rentabilidad de Proyectos	1	2	3
6	Evaluación de Proyectos de Inversión	1	2	3
7	Elaboración de Proyectos de Ingeniería	3	5	1
8	Evaluación de Costos y Tiempos de Producción	1	2	2
9	Factibilidad Económica de Proyectos	1	2	3
10	Elaboración de Documentación Técnica	2	3	4
11	Metodologías Ágiles en Diseño de Soluciones	1	2	2
12	Uso de Herramientas de Emprendimiento	1	1	4
13	Evaluación de Oportunidades y Prototipos	1	2	3
14	Análisis de Problemas Operativos	0	4	4
15	Validación Comercial de Prototipos	1	1	2
16	Estimación de Mercado y Barreras	1	1	3
17	Construcción de Plan de Negocios	1	1	4
18	Diseño de Modelos de Negocio	1	1	3
19	Prospección de Necesidades Sociales	1	1	1
20	Validación de Tracción de Proyecto	1	1	5
21	Trabajo en Equipo en Proyectos Tecnológicos	3	3	4
22	Redacción de Informes de Proyecto	2	3	1
23	Presentación de Soluciones Técnicas	3	6	1
24	Creación y Gestión de Equipos	1	5	1
25	Orientación al Propósito Colectivo	1	5	5
26	Comunicación Efectiva de Proyectos	3	6	4
27	Reevaluación de Propuesta de Valor	1	1	1
28	Testeo de Soluciones con Maquetas	1	1	2
29	Evaluación Crítica de Acciones	3	6	2
30	Identificación y Solución de Problemas de Equipo	3	6	5
31	Evaluación Económica y Social de Productos	1	2	3
32	Validación de Producto Mínimo Viable	1	1	5*
33	Pitch y Exposiciones Orales	3	6	3
34	Exposición de Proyectos de Inversión	3	5	2
35	Producción de Reportes Técnicos	2	3	4
36	Presentación con Recursos Audiovisuales	1	2	4
37	Evaluación Integral de Soluciones	3	6	5

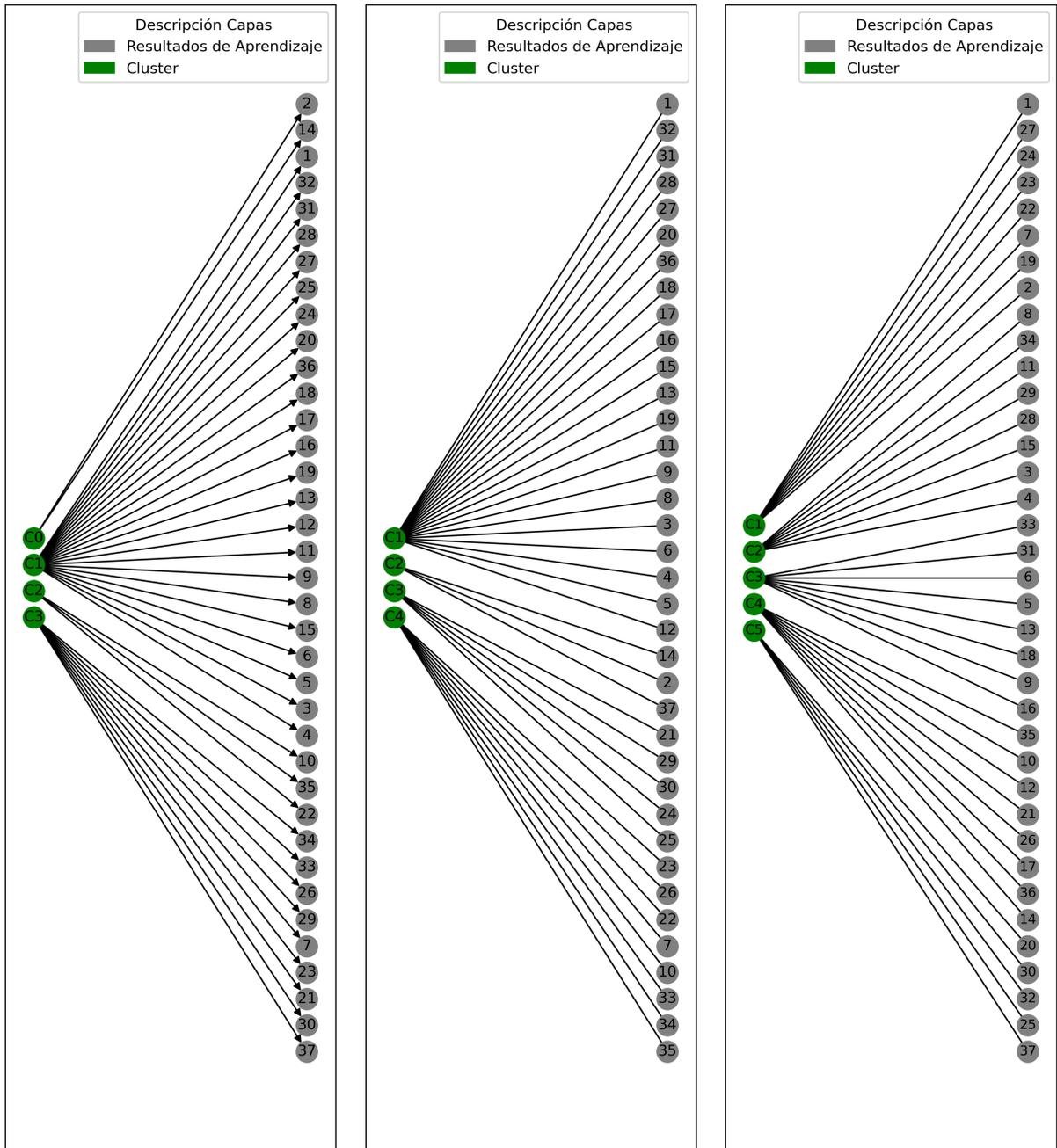


Figura 5.13: Visualización de los grupos creados por los diferentes métodos de agrupación. Izquierda: Clústeres creados por K-Means. Centro: Agrupación jerárquica. Derecha: Agrupación con GPT-4o.

Capítulo 6

Discusión

6.1. Análisis del currículum de Ingeniería Civil Eléctrica

Las Figuras 5.1 y 5.2 muestran la relación entre los resultados de aprendizaje que se abordan en los cursos con las competencias específicas y generales a las que tributan, respectivamente. De ambas figuras, se destaca el mapeo realizado para el curso EL6201, ya que abarca un total de 9 resultados de aprendizaje relacionados a competencias específicas y 18 resultados de aprendizaje relacionados a competencias genéricas.

Al hacer un análisis consecuente con la descripción de los resultados de aprendizaje de la Tabla A.1 y los cursos considerados en este estudio, se obtiene lo siguiente:

- EL3105: se abarcan resultados de aprendizaje que están en general enfocados a nivel macro de soluciones.
- EL4204: se considera un curso enfocado a lo teórico respecto a evaluación de proyectos e indicadores económicos de proyectos.
- EL6101: se puede observar que este curso profundiza en relación al proceso de innovación, abarcando aspectos como el desarrollo de productos mínimos viables, documentación técnica y trabajo en equipo.
- EL6201: es el curso con más resultados de aprendizaje asociados. Destacan varios resultados de aprendizaje asociados al trabajo en equipo, y a la gestión del proyecto desde la evaluación de oportunidades y la iteración y validación con usuarios.

6.2. Preparación de los datos

6.2.1. Comparación de modelos de clasificación

El objetivo de la clasificación en esta etapa, es identificar las ofertas de trabajo que no son relevantes para la investigación, es decir, identificar datos que ensucian la base de datos. A partir de esto, los falsos positivos tienen un efecto mucho más perjudicial para el trabajo que los falsos negativos, puesto que significa mantener en el conjunto de datos ofertas que aportan información que puede confundir al modelo en las etapas siguientes. Por otro lado,

un falso negativo significa que se pierden ofertas de empleo, lo que no resulta un problema si el porcentaje de pérdida no es significativo.

Considerando lo descrito en el párrafo anterior, al observar la Figura 5.6, es posible notar que la clasificación mediante GPT-4o no obtiene falsos negativos. Por otro lado, la clasificación del modelo Bart obtiene 9 falsos positivos (30 % de los datos), resultando en una clara desventaja para este último modelo.

Al considerar además, la información de la Tabla 5.4, se puede observar que el modelo GPT-4o pierde 5 observaciones que originalmente debieron haber sido clasificadas como “Inteligencia Artificial”, equivalente al 35.7 % de pérdida de datos.

Por el lado de la clasificación *Zero-Shot*, si bien el número de clasificaciones bajo la etiqueta de “Inteligencia Artificial” (13), se acerca mucho más al total real (14), al considerar la Figura 5.6, se puede verificar que de los 13 valores clasificados en la categoría positiva, sólo 4 están correctamente clasificados. Esto equivale a un 71.4 % de pérdida de datos.

Las métricas de la Tabla 5.5, se condicen con el análisis hecho, puesto que el modelo GPT-4o obtiene una precisión igual a 1, distando del modelo Bart que obtiene un valor de 0.308. Esta métrica toma especial relevancia cuando los resultados se ven más perjudicados con los falsos positivos, como es el caso de este estudio.

Todas las métricas estudiadas obtienen resultados favorables para el modelo GPT-4o. Considerando además, el análisis hecho previamente, se seleccionó dicho modelo para clasificar el conjunto de datos completo y así poder filtrar los registros no relevantes.

Además de las métricas estudiadas, es importante considerar que en este experimento fueron clasificadas en la categoría “Otros” un total de 21 ofertas de empleo, lo que se traduce en un 70 % de los datos.

6.2.2. Eliminación de registros no relevantes

Según los resultados de la subsección anterior, se procedió con el filtrado de la base de datos completa, utilizando 9 etiquetas.

La Tabla 5.6 muestra algunas métricas relevantes para la evaluación del modelo. Por un lado, se obtiene que al modelo le toma cerca de 73 minutos procesar un total de 4.499 ofertas de empleo. El procesamiento de las ofertas de empleo utilizó un total de 3.100.722 *tokens*, marcando fuertemente la diferencia entre los de contexto (2.938.811) y los generados como respuesta (161.911). Esta diferencia se explica por el tipo de entrada en el mensaje de contexto e indicación, puesto que cada vez que se ejecuta la consulta se ingresan todas las etiquetas y las indicaciones explicitadas en la metodología. Por otro lado, la respuesta solicitada fue bastante restrictiva. Se solicitó una respuesta con formato de lista, especificando además, no agregar información adicional en la respuesta. A pesar de esto, el modelo no siempre entregó solamente la lista solicitada, por lo que fue necesario extraerla de la respuesta mediante procesamiento de datos. Esto explica también la cantidad de *tokens* generados.

La cantidad de *tokens* totales se traduce en un costo igual a 17.12 USD.

La Tabla 5.7, contiene una comparación de los resultados iniciales y la cantidad ofertas asignadas a cada etiqueta por el modelo GPT-4o. Considerando esto, se observa que la mayor diferencia entre cantidades clasificadas, se obtiene para la categoría “Ingeniero de datos” puesto que el modelo identificó tan sólo 55 ofertas de empleo en esta categoría, alcanzando un porcentaje equivalente al 6.0% de la cantidad de ofertas encontradas inicialmente para la búsqueda “Ingeniero de datos”.

Con la Figura 5.7, se visualizan las cantidades por etiquetas, permitiendo dimensionar las cantidades clasificadas. En particular, se destaca que la cantidad de ofertas de empleo clasificadas como “Otros” alcanza un total de 2.019, equivalente al 46% de los datos ingresados inicialmente.

Finalmente, la base de datos quedó compuesta por 2.468 registros.

6.3. Etiquetado de la base de datos

Continuando con el análisis del proceso de etiquetado de la base de datos utilizando los 37 resultados de aprendizaje detallados en la Tabla 5.2, se obtienen las características de la ejecución del programa de la Tabla 5.8. Se puede observar que el tiempo de ejecución fue igual a 53 minutos con 37 segundos. Al comprar la cantidad de *tokens* totales utilizados en esta ejecución con los resultados de la etapa de filtrado (Tabla 5.6), se obtiene una diferencia de aproximadamente 1 millón de *tokens*. Esta diferencia, se explica puesto que la cantidad de ofertas de empleo procesadas fue notablemente menor (2.468 registros), dado que en este punto ya habían sido eliminadas las ofertas no relevantes para el estudio.

6.3.1. Análisis preliminar de los resultados

Los resultados de la Tabla 5.9 y la Figura 5.8, muestran la cantidad de ofertas de trabajo en las que fue identificado cada uno de los 37 resultados de aprendizaje. Se observa que los 5 resultados de aprendizaje con mayor relación con lo buscado en el mercado laboral, según este análisis, son:

1. Trabajo en Equipo en Proyectos Tecnológicos (21): 2.413 apariciones.
2. Análisis de Problemas Operativos (14): 2.365 apariciones.
3. Identificación de Problemas y Oportunidades (2): 2.337 apariciones.
4. Producción de Reportes Técnicos (35): 2.100 apariciones.
5. Elaboración de Documentación Técnica (10): 2.047 apariciones.

Por otro lado, los 5 resultados de aprendizaje con menor relación con lo buscado en el mercado laboral son:

1. Uso de Herramientas de Emprendimiento (12): 14 apariciones.

2. Diseño e Implementación de Prototipos Innovadores (3): 30 apariciones.
3. Testeo de Soluciones con Maquetas (28): 43 apariciones.
4. Validación Comercial de Prototipos (15): 48 apariciones.
5. Diseño de Modelos de Negocio (18): 58 apariciones.

Para analizar estos resultados, se considera la Tabla 5.2, que contiene la agrupación de resultados de aprendizaje en subcompetencias, competencias y grupos de competencias.

Por el lado de los resultados de aprendizaje menos relacionados, se detectan tres subcompetencias y dos competencias, es decir, estos resultados de aprendizaje se encuentran relacionados entre sí. En particular las subcompetencias implicadas son:

1. Herramientas y Métodos de Emprendimiento (RA 12 y 18) → Gestión de Proyectos y Emprendimiento.
2. Diseño de Prototipos (RA 3 y 28) → Diseño y Prototipado de Soluciones.
3. Evaluación de Prototipos y Productos (RA 15) → Diseño y Prototipado de Soluciones.

Se puede observar que al subir un nivel más en la jerarquía, las competencias implicadas son dos: Gestión de Proyectos y Emprendimiento, y Diseño y Prototipado de Soluciones. Esto indica una mayor relación entre los resultados de aprendizaje mencionados.

Al considerar además, las descripciones de estos resultados de aprendizaje, se reafirma esta relación entre resultados de aprendizaje. En particular, los resultados de aprendizaje 3, 15 y 28 apuntan al desarrollo de prototipos y *mockups* para validación de soluciones. Es razonable que esta competencia no sea buscada explícitamente en ofertas de empleo dirigidas a Ingenieros Eléctricos, ya que es un carrera que suele ser más técnica que de negocios. Si bien los resultados de aprendizaje 12 y 18 no están directamente relacionados con el primer grupo mencionado, ambos apuntan a la gestión de proyectos y emprendimiento, por lo que la justificación de estar dentro de los menos buscados recae en la misma anterior.

Analizando los resultados de aprendizaje más buscados en el estudio, la agrupación en niveles más altos de la jerarquía no es directa como para los menos buscados. El resultados de aprendizaje Trabajo en Equipo en Proyectos Tecnológicos aparece como el más demandado. El trabajo en equipo se considera una competencia fundamental en el área de la ingeniería, por lo que es un resultado esperado.

Aunque los resultados de aprendizaje en el segundo y tercer lugar, no se relacionan directamente según la agrupación de competencias y subcompetencias, ambos apunta a la identificación de problemas, pero en distintas etapas de proyecto, por lo que sí existe relación entre ellas. Ambas se encuentran dentro de las más demandadas puesto que dentro de los trabajos fundamentales de un ingeniero se encuentra la identificación y resolución de problemas en distintos ámbitos.

Finalmente, el resultado de aprendizaje número 10 se relaciona directamente a través de las competencias y subcompetencias identificadas con el resultado de aprendizaje 35. Ambos

apuntan a la creación de documentación técnica, por lo que se justifica que aparezcan juntos, con similar cantidad de búsquedas.

Respecto a la Tabla 5.10, esta presenta características relevantes para comprender el resultado de etiquetado de la base de datos. En promedio, cada oferta de empleo fue asociada a 13 resultados de aprendizaje. Dentro de las 2.468 ofertas de empleo, la oferta que alcanzó la mayor cantidad de resultados de aprendizaje identificados fue única, con 31 resultados de aprendizaje. Por otro lado, la cantidad mínima de resultados de aprendizajes identificados fue 1, obteniendo 4 ofertas de trabajo con este resultado.

La Figura 5.9, entrega una visualización de la cantidad de ofertas de empleo que obtuvieron x resultados de aprendizaje identificados (con x un entero $\in [1, 31]$). Se observa una distribución normal levemente desplazada, centrada entre 12 y 13 resultados de aprendizaje, lo que coincide con el promedio calculado en la Tabla 5.10.

6.4. Construcción de la propuesta

El método del codo se utiliza para determinar el número óptimo de clústeres a formar con K-Means. En este caso, la conclusión obtenida de la Figura 5.10 fue óptimo igual a 4, puesto que a partir de ese punto la pendiente de caída de la inercia se vuelve menor.

Sobre la Figura 5.11, bajo simple inspección se observan clústeres desbalanceados en cuanto a cantidad de resultados de aprendizaje incluidos en cada uno. Más adelante, se hará un análisis completo respecto a los resultados obtenidos y el sentido de ellos.

La Figura 5.12, contiene el resultado de la agrupación jerárquica utilizando el método *linked()*. El corte de distancia se realizó en el valor 40, entregando así 6 clústeres formados. Se observa una clara división entre los dos primeros clústeres formados (naranja y verde), en relación a los formados por el lado derecho de la figura. Se puede ver que los clústeres rojo, morado, café y rosado tienen mayor similitud entre ellos que con los clusters naranja y verde. Así, a pesar de que la distancia de corte fue fijada en 40, es posible unificar los clústeres de la derecha, pensando en la posible construcción de cursos para la carrera de Ingeniería Civil Eléctrica.

La Tabla 5.11, contiene parte del resultado de la respuesta entregada por GPT-4o ante la solicitud de agrupación de cursos. En particular, en la tabla se añaden los nombres de curso entregados por el modelo, acompañado del enfoque detectado para cada uno.

Del resultado, se observa inmediatamente que las agrupaciones obtenidas tienen un sentido de formación, apuntando cada uno a abordar diferentes temáticas relacionadas a la formación en innovación. Además, entre los cursos 4 y 5 de la Tabla 5.11, se aprecia cierta similitud, detectando una posible reducción de resultados.

6.5. Análisis final y consolidación de la propuesta

Considerando las Tablas 5.12 y A.1, y la Figura 5.13, es posible analizar los resultados desde la perspectiva del objetivo de este trabajo de título: proponer opciones de mejora en el plan de formación en innovación de la Carrera de Ingeniería Civil Eléctrica. Este análisis se realizará en 4 partes:

- Análisis de resultados entregados por K-Means.
- Análisis de resultados entregados por la agrupación jerárquica.
- Análisis de resultados entregados por GPT-4o.
- Consolidación de la propuesta.

6.5.1. K-Means

En la Tabla 5.12, la columna K-Means contiene el clúster asignado a cada resultado de aprendizaje por el algoritmo. La construcción de los cursos propuesta por este algoritmo es de la siguiente forma:

1. **Curso 1 (clúster 0):** 2 Resultados de Aprendizaje.

- | | |
|---|---------------------------------------|
| 2. Identificación de Problemas y Oportunidades. | 14. Análisis de Problemas Operativos. |
|---|---------------------------------------|

2. **Curso 2 (clúster 1):** 23 Resultados de Aprendizaje.

- | | |
|--|---|
| 1. Análisis de Desafío Innovación. | 16. Estimación de Mercado y Barreras. |
| 3. Diseño de Prototipos Innovadores. | 17. Construcción de Plan de Negocios. |
| 4. Formulación de Proyectos de Inversión. | 18. Diseño de Modelos de Negocio. |
| 5. Cálculo de Rentabilidad de Proyectos. | 19. Prospección de Necesidades Sociales. |
| 6. Evaluación de Proyectos de Inversión. | 20. Validación de Tracción de Proyecto. |
| 8. Evaluación de Costos y Tiempos de Producción. | 24. Creación y Gestión de Equipos. |
| 9. Factibilidad Económica de Proyectos. | 25. Orientación al Propósito Colectivo. |
| 11. Metodologías Ágiles en Diseño de Soluciones. | 27. Reevaluación de Propuesta de Valor. |
| 12. Uso de Herramientas de Emprendimiento. | 28. Testeo de Soluciones con Maquetas. |
| 13. Evaluación de Oportunidades y Prototipos. | 31. Evaluación Económica y Social de Productos. |
| 15. Validación Comercial de Prototipos. | 32. Validación de Producto Mínimo Viable. |
| | 36. Presentación con Recursos Audiovisuales. |

3. **Curso 3 (clúster 2):** 3 Resultados de Aprendizaje.

- | | |
|---|--|
| 10. Elaboración de Documentación Técnica. | 22. Redacción de Informes de Proyecto. |
| | 35. Producción de Reportes Técnicos. |

4. **Curso 4 (clúster 3):** 9 Resultados de Aprendizaje.

- | | |
|--|---|
| 7. Elaboración de Proyectos de Ingeniería. | 30. Identificación y Solución de Problemas de Equipo. |
| 21. Trabajo en Equipo en Proyectos Tecnológicos. | 33. Pitch y Exposiciones Orales. |
| 23. Presentación de Soluciones Técnicas. | 34. Exposición de Proyectos de Inversión. |
| 26. Comunicación Efectiva de Proyectos. | 37. Evaluación Integral de Soluciones. |
| 29. Evaluación Crítica de Acciones. | |

Al revisar los resultados de aprendizaje incluidos en cada uno de los clústeres, lo primero que se debe comentar es el desbalanceo entre la cantidad de resultados de aprendizaje que compone cada grupo. Resulta llamativo que el primer curso formado contenga 2 resultados de aprendizaje en comparación con el segundo curso formado, que contiene un total de 23 resultados de aprendizaje.

Recordando las Figuras 5.1 y 5.2, el curso EL6201 está compuesto por un total del 19 resultados de aprendizaje distintos, por lo que el número alcanzado por el segundo curso formado mediante K-Means, no se encontraría alejado de lo que se podría realmente aplicar en un curso de la carrera. Al analizar a nivel de competencias de la Tabla 5.2, se observa que dentro de los 23 resultados de aprendizaje son abordadas todas las competencias. Al bajar a nivel de subcompetencias en esta propuesta de curso, se abordarían 8 de las 10 subcompetencias definidas. Esto habla de un curso poco especializado que aborda de forma superficial cada tema. Sin embargo, al analizar el curso EL6201, este considera 5 de las 5 competencias, y 9 de las 10 subcompetencias, indicando que el resultado de agrupación sigue siendo viable. Finalmente, es importante notar que el Curso 2, creado con K-Means, y el curso EL6201, coinciden en 14 resultados de aprendizaje.

El Curso 3, se encuentra compuesto de 3 resultados de aprendizaje. A pesar de ser una baja cantidad en comparación con el Curso 2, resulta relevante la relación entre ellos, ya que todos corresponden a la competencia de Comunicación y Documentación Técnica y a la subcompetencia de Documentación Técnica. Se destaca como un resultado relevante, puesto que entre los datos ingresados para la agrupación con este método no se encontraba la organización a nivel de grupos de competencias, competencias y subcompetencias. Con estos resultados, el Curso 3 consideraría de un abordaje más especializado en el tema de la documentación técnica.

Al analizar el Curso 4 y comparar los resultados con la Tabla 5.2, a nivel de competencias se consideran 3: Diseño y Prototipado (7 y 37), Comunicación Documentación (22, 23, 26, 33 y 34) y, finalmente, Trabajo en Equipo y Liderazgo (21, 29 y 30). Por un lado, nuevamente se detectan coincidencias en el resultado a nivel de competencias, ya que se agrupan por temáticas similares. Por otro, al considerar los resultados de aprendizaje seleccionados y sus

subcompetencias, este curso abordaría fuertemente el trabajo en equipo y las habilidades de presentación de prototipos.

Finalmente, para los resultados de aprendizaje entregados en el Curso 1, no es posible encontrar una relación directa en los niveles más bajos de organización de la Tabla 5.2, ya que solo coinciden en el Grupo de Competencias. A pesar de esto, la descripción de ambos resultados de aprendizaje apunta a la identificación de problemas, pero en diferentes etapas del proyecto. Es importante comentar que el número de resultados de aprendizaje identificados podría suponer un problema, ya que es demasiado bajo para la construcción de un curso.

6.5.2. Agrupación Jerárquica

En la columna Jerárquico de la Tabla 5.12, se muestran los clústeres asignados a cada resultados de aprendizaje por este método, obteniendo la siguiente propuesta:

1. **Curso 1 (clúster 1):** 11 Resultados de Aprendizaje.

- | | |
|--|---|
| 1. Análisis de Desafío Innovación. | 19. Prospección de Necesidades Sociales. |
| 12. Uso de Herramientas de Emprendimiento. | 20. Validación de Tracción de Proyecto. |
| 15. Validación Comercial de Prototipos. | 27. Reevaluación de Propuesta de Valor. |
| 16. Estimación de Mercado y Barreras. | 28. Testeo de Soluciones con Maquetas. |
| 17. Construcción de Plan de Negocios. | 32. Validación de Producto Mínimo Viable. |
| 18. Diseño de Modelos de Negocio. | |

2. **Curso 2 (clúster 2):** 10 Resultados de Aprendizaje.

- | | |
|--|--|
| 3. Diseño de Prototipos Innovadores. | 11. Metodologías Ágiles en Diseño de Soluciones. |
| 4. Formulación de Proyectos de Inversión. | 13. Evaluación de Oportunidades y Prototipos. |
| 5. Cálculo de Rentabilidad de Proyectos. | 31. Evaluación Económica y Social de Productos. |
| 6. Evaluación de Proyectos de Inversión. | 36. Presentación con Recursos Audiovisuales. |
| 8. Evaluación de Costos y Tiempos de Producción. | |
| 9. Factibilidad Económica de Proyectos. | |

3. **Curso 3 (clúster 3):** 4 Resultados de Aprendizaje.

- | | |
|---|--|
| 10. Elaboración de Documentación Técnica. | nológicos. |
| 21. Trabajo en Equipo en Proyectos Tec- | 22. Redacción de Informes de Proyecto. |
| | 35. Producción de Reportes Técnicos. |

4. **Curso 4 (clúster 4):** 2 Resultados de Aprendizaje.

- | | |
|---|---------------------------------------|
| 2. Identificación de Problemas y Oportunidades. | 14. Análisis de Problemas Operativos. |
|---|---------------------------------------|

5. **Curso 5 (clúster 5):** 3 Resultados de Aprendizaje.

- | | |
|--|---|
| 7. Elaboración de Proyectos de Ingeniería. | 25. Orientación al Propósito Colectivo. |
| 24. Creación y Gestión de Equipos. | 34. Exposición de Proyectos de Inversión. |

6. **Curso 6 (clúster 6):** 6 Resultados de Aprendizaje.

- | | |
|--|---|
| 23. Presentación de Soluciones Técnicas. | 30. Identificación y Solución de Problemas de Equipo. |
| 26. Comunicación Efectiva de Proyectos. | 33. Pitch y Exposiciones Orales. |
| 29. Evaluación Crítica de Acciones. | 37. Evaluación Integral de Soluciones. |

Se realizó la agrupación jerárquica en $d=40$, lo cual generó un total de 6 clústeres, por ende, 6 posibles propuestas de cursos.

Al analizar el Curso 1 en relación a la Tabla 5.2, se observa relación a nivel de competencias. Los 11 resultados de aprendizaje seleccionados abarcan un total de 3 competencias distintas: Identificación y Análisis de Problemas (1, 19, 20 y 27), Gestión de Proyectos y Emprendimiento (12, 16, 17 y 18) y, finalmente, Diseño y Prototipado (15, 28 y 32). Además de la relación directa que tienen los resultados de aprendizaje a través del nivel de competencias, es importante notar que 10 de los 11 resultados de aprendizaje coinciden con el curso EL6201, lo que es un resultado relevante de agrupación, puesto que esta información no fue incluida en los datos ingresado al modelo de agrupación. El resultado propone un curso enfocado en las etapas previas al desarrollo del proyecto, considerando fuertemente la identificación de dolores para la propuesta de una solución. Además, considera la planificación de un proyecto.

Al igual que en el Curso 1, en el Curso 2 se obtiene relación directa a nivel de competencias. En este caso, para los 10 resultados de aprendizaje seleccionados, 9 de ellos se dividen entre 2 competencias: Diseño y Prototipado de soluciones (3, 4, 5, 6 y 31) y Gestión de Proyectos y Emprendimiento (8, 9, 11 y 13). Se añade una competencia adicional, Comunicación y Documentación, con un resultado de aprendizaje asociado. En este grupo, es posible además encontrar relaciones a nivel de subcompetencias, con 8 de los resultados distribuidos en tan solo 3 subcompetencias (Diseño de Prototipos, Evaluación de Prototipos y Productos, y Planificación y Evaluación de Proyectos). La orientación dada por los resultados para este curso, se enfoca en el el diseño de prototipos y evaluación de indicadores de mercado para ellos. Considera la planificación de un proyecto mediante herramientas para la innovación.

Respecto a la relación de los Cursos 1 y 2 de la agrupación jerárquica, y el resultado generado con K-Means, se observa que el 100% de los resultados de aprendizaje de ambos cursos pertenecen al Curso 2 de K-Means, es decir, los Cursos 1 y 2 de la agrupación jerárquica son subconjuntos del Curso 2 de K-Means.

El Curso 3 generado por la agrupación jerárquica, es similar al Curso 3 generado por K-Means. La única diferencia se encuentra en la adición de un resultado de aprendizaje (21. Trabajo en Equipo en Proyectos Tecnológicos) al clúster en el resultado para agrupación jerárquica.

El Curso 4 construido mediante agrupación jerárquica resulta idéntico al Curso 1 de los resultados obtenidos con K-Means.

Para el quinto clúster formado no tiene relaciones directas significantes a niveles de competencias o subcompetencias, pero se puede caracterizar como un grupo de resultados de aprendizaje orientados al trabajo en equipo y la exposición de resultados de forma efectiva. A nivel de competencias y subcompetencias abordadas, se observa un resultado similar al Curso 4 de K-Means.

Finalmente el Curso 6, considera un total de 3 competencias (Comunicación y Documentación, Trabajo en Equipo y Liderazgo y Diseño y Prototipado de soluciones). Es posible caracterizar el grupo nuevamente como un curso orientado al trabajo en equipo y comunicación efectiva de resultados. La relación con el Curso 4 de K-Means se mantiene, notando además, que la combinación de los Cursos 5 y 6 es equivalente a 8 de los 9 resultados de aprendizaje del Curso 4 de K-Means.

Tras el análisis anterior, es directo pensar en agrupar algunos de los clústeres con menor cantidad de resultados, especialmente al pensar en el ejemplo de la relación entre los cursos 5 y 6, puesto que considerarían temáticas similares. El proceso de agrupación también respalda esto. Al observar la Figura 5.12, específicamente los clústeres 5 (café) y 6 (rosado), se observa que el siguiente nivel de agrupación jerárquica considera la unión de estos dos grupos en una distancia cercana ($d = 50$, aproximadamente).

6.5.3. Agrupación con GPT-4o

El resultado obtenido a partir de la agrupación con GPT-4o fue distinto a los dos métodos de agrupación tradicionales revisados anteriormente. La respuesta del modelo, por su naturaleza, fue mucho más completa ya que es capaz de entregar justificaciones de las decisiones y notas adicionales respecto a la respuesta entregada.

Otro punto relevante que se debe considerar es que la información que se le entregó al modelo de OpenAI fue mucho más completa que para los procesos de agrupación, puesto que se ingresó toda la información de las relaciones en niveles de agrupación (grupos de competencias, competencias, subcompetencias y resultados de aprendizaje), descripciones de cada resultado de aprendizaje, y los pesos obtenidos a partir del análisis de las ofertas de empleo. Adicionalmente, se entregaron guías específicas para la construcción de los cursos, como por ejemplo, la necesidad de que los resultados de aprendizaje fuesen organizados en 3 a 5 grupos, para poder generar una propuesta viable para la carrera de Ingeniería Civil Eléctrica (para estudio se consideraron los 4 cursos obligatorios actuales del plan curricular).

La respuesta entregada por el modelo, consideró un total de 5 cursos. La Tabla 5.11, muestra el nombre de cada curso y la descripción del enfoque u objetivo asignado por GPT-4o. Los

enfoque de los 3 primeros cursos definen objetivos claros para cada uno de ellos. La definición y separación de objetivos no es tan directa entre los cursos 4 y 5, ya que ambos consideran el monitoreo en el desarrollo de proyectos para lograr mejores resultados.

Respecto al detalle de resultados de aprendizaje seleccionados por el modelo para cada curso, reflejado en la Tabla 5.12 en la columna GPT-4o, se obtiene lo siguiente:

1. **Curso 1 (clúster 1):** 8 Resultados de Aprendizaje.

- | | |
|---|--|
| 1. Análisis de Desafío Innovación. | 22. Redacción de Informes de Proyecto. |
| 2. Identificación de Problemas y Oportunidades. | 23. Presentación de Soluciones Técnicas. |
| 7. Elaboración de Proyectos de Ingeniería. | 24. Creación y Gestión de Equipos. |
| 19. Prospección de Necesidades Sociales. | 27. Reevaluación de Propuesta de Valor. |

2. **Curso 2 (clúster 2):** 8 Resultados de Aprendizaje.

- | | |
|--|---|
| 3. Diseño de Prototipos Innovadores. | luciones. |
| 4. Formulación de Proyectos de Inversión. | 15. Validación Comercial de Prototipos. |
| 8. Evaluación de Costos y Tiempos de Producción. | 28. Testeo de Soluciones con Maquetas. |
| 11. Metodologías Ágiles en Diseño de So- | 29. Evaluación Crítica de Acciones. |
| | 34. Exposición de Proyectos de Inversión. |

3. **Curso 3 (clúster 3):** 8 Resultados de Aprendizaje.

- | | |
|---|---|
| 5. Cálculo de Rentabilidad de Proyectos. | 16. Estimación de Mercado y Barreras. |
| 6. Evaluación de Proyectos de Inversión. | 18. Diseño de Modelos de Negocio. |
| 9. Factibilidad Económica de Proyectos. | 31. Evaluación Económica y Social de Productos. |
| 13. Evaluación de Oportunidades y Prototipos. | 33. Pitch y Exposiciones Orales. |

4. **Curso 4 (clúster 4):** 8 Resultados de Aprendizaje.

- | | |
|--|--|
| 10. Elaboración de Documentación Técnica. | 21. Trabajo en Equipo en Proyectos Tecnológicos. |
| 12. Uso de Herramientas de Emprendimiento. | 26. Comunicación Efectiva de Proyectos. |
| 14. Análisis de Problemas Operativos. | 35. Producción de Reportes Técnicos. |
| 17. Construcción de Plan de Negocios. | 36. Presentación con Recursos Audiovisuales. |

5. **Curso 5 (clúster 5):** 5 Resultados de Aprendizaje.

- | | |
|---|---|
| 20. Validación de Tracción de Proyecto. | 32. Validación de Producto Mínimo Viable. |
| 25. Orientación al Propósito Colectivo. | |
| 30. Identificación y Solución de Problemas de Equipo. | 37. Evaluación Integral de Soluciones. |

Es importante mencionar que la cantidad de resultados de aprendizaje por curso se encuentra mucho más equilibrada en este resultado en comparación a los obtenidos a partir de la agrupación con K-Means y la agrupación jerárquica. Este equilibrio en la distribución de la cantidad de resultados de aprendizaje por curso se presenta de forma visual en la Figura 5.13, donde la red del lado derecho (correspondiente a GPT-4o) muestra una distribución similar en la cantidad entre cursos.

Otro punto a destacar, es que los resultados de aprendizaje orientados a comunicación y trabajo en equipo fueron distribuidos entre todos los cursos formados, puesto que fue una indicación explícita ingresada en el *prompt*.

El Curso 1 propuesto por GPT-4o recibe el nombre “Análisis e Identificación de Problemas y Oportunidades Innovadoras”. Los resultados de aprendizaje considerados en este grupo están fuertemente orientados a la identificación y definición de problemas, y diseño de prototipos en base a lo levantado inicialmente. Considera además, el desarrollo de la comunicación escrita y oral y de trabajo en equipo. El enfoque propuesto por GPT-4o en la Tabla 5.11 para este curso, coincide con lo descrito anteriormente.

Al comparar el Curso 1 propuesto por GPT-4o, se detecta similitud con el curso EL3501, ya que el enfoque de ambos cursos es similar. La principal diferencia se encuentra en que el curso que actualmente pertenece a la malla de Ingeniería Civil Eléctrica, no considera resultados de aprendizaje relacionados al trabajo en equipo, a diferencia del Curso 1 propuesto por GPT-4o.

El Curso 2 propuesto por la inteligencia de OpenAI (Diseño y Prototipado de Soluciones), considera dos competencias principales: Diseño y prototipado de soluciones, y Gestión de proyectos y emprendimiento. Todos los resultados de aprendizaje considerados se orientan al diseño de prototipos, evaluación de tiempos y económica del proyecto, y validación de prototipos. Además, se incluyen componentes de trabajo en equipo y comunicación oral y escrita. El objetivo dado por GPT-4o coincide plenamente con los resultados de aprendizaje seleccionados.

Al comprar el Curso 2 con los cursos actuales de la carrera se detecta una fuerte similitud con el curso EL6101, especialmente en el enfoque propuesto. La principal diferencia de estos cursos se encuentra en la inclusión de la validación comercial tanto del dolor identificado, como del prototipo construido.

Al analizar el tercer curso propuesto en este resultado (Evaluación y Gestión de Proyectos Innovadores), la mayoría de resultados de aprendizaje están asociados a la gestión de proyectos y emprendimiento, y a la obtención de indicadores económicos mediante evaluaciones sociales y económicas de proyectos. Este curso tiene una fuerte relación con el curso EL4204. Si bien los resultados de aprendizaje asignados en su mayoría corresponden a otros cursos,

las descripciones de ellos apuntan al mismo objetivo de evaluación económica de proyectos. Entre los resultados de aprendizaje propuestos por GPT-4o se encuentra el diseño de un modelo de negocio, lo cual sería una diferencia entre el diseño actual del curso EL4204.

EL Curso 4 se enfoca en la implementación de proyectos, priorizando el punto de vista de la gestión del proyecto, presentación de resultados y trabajo en equipo. Tiene una fuerte similitud con el curso EL6201, ya que de los 8 resultados de aprendizaje asignados por GPT-4o en este curso, 5 de ellos corresponden a resultados de aprendizaje del curso EL6201. La principal diferencia se encuentra en que la propuesta hecha por GPT-4o considera mucho más relevante para el curso la comunicación oral y escrita, incluyendo la elaboración de reportes técnicos y, además, deja fuera la identificación de problemas y oportunidades.

Finalmente, el Curso 5 considera un enfoque similar al del Curso 4. Incluyendo en las competencias abordadas la identificación y análisis de problemas considerando la tracción del proyecto. Se repite la relación con el curso EL6201, ya que de los 5 resultados de aprendizaje asignados a Curso 5, 4 pertenecen al curso EL6201. La diferencia entre estos cursos es que la propuesta de GPT-4o deja fuera de las consideraciones la gestión, evaluación y planificación de proyectos.

Al analizar los Cursos 4 y 5 resulta directo realizar el análisis de ambos cursos como uno solo. Así, se obtiene una marcada similitud al curso EL6201. Estos cursos se encuentran enfocados a la planificación y gestión de proyectos, así como a la validación de tracción y trabajo en equipo.

6.5.4. Consolidación de la Propuesta

Tras haber analizado en detalle cada uno de los resultados obtenidos, se puede detectar diferencias en las formas de agrupar de ambos métodos. La agrupación (K-Means y jerárquica) agrupa resultados de aprendizaje según la similitud entre ellos, obteniendo que los resultados de aprendizaje que conducen a competencias similares son agrupados. Por otro lado, el resultado de GPT-4o consigue agrupar los resultados de una forma que podría resultar más secuencial en el desarrollo de cada una de las competencias.

Al considerar además, el plan de formación actual de la carrera, se pueden proponer los siguientes puntos, enfocados al cierre de brechas entre el plan actual y lo detectado en esta investigación:

1. EL3105: Seminario de Innovación Tecnológica.

Actualmente, este curso es considerado introductorio a los cursos de proyecto en Ingeniería Eléctrica. Busca que los estudiantes conozcan las opciones de desarrollo en la carrera, incluyendo la innovación como una de ellas. La principal diferencia identificada fue la falta de resultados de aprendizaje relacionados al trabajo en equipo, por lo que siguiendo los resultados obtenidos y considerando el cierre de brechas, se propone la inclusión de este en el curso desde el punto de vista de colaboración y gestión de equipos. Esta propuesta se ve justificada además, en el análisis del mercado laboral, puesto que el trabajo en equipo lidera el ranking de habilidades requeridas.

2. **EL4204: Formulación y Evaluación de Proyectos.**

Tras revisar los resultados de aprendizaje y, según el análisis realizado previamente, se detecta una oportunidad de mejora en la inclusión de un resultado de aprendizaje al diseño de un modelo de negocio de baja resolución, puesto que para realizar la evaluación económica de proyectos es necesario conocer la forma de ingresos para estimaciones más acertadas.

3. **EL6101: Taller de Proyectos Tecnológicos**

Este curso tiene como objetivo el desarrollo de un producto a partir de un estudio de ingeniería. El foco en este curso se encuentra en el desarrollo de un proyecto real de ingeniería. Dentro de las posibles mejoras identificadas se encuentra la inclusión de un resultado de aprendizaje enfocado en la iteración y validación el usuario y/o cliente.

4. **EL6201: Taller de Emprendimiento e Innovación.**

El curso EL6201 tiene como foco la identificación de un dolor (ya sea comercial o social) con el objetivo de desarrollar soluciones innovadoras. Este curso se centra en el desarrollo de las ideas y la evaluación económica del proyecto, y no en la creación de prototipos. Según los resultados mencionados previamente, se propone la inclusión de un resultado de aprendizaje relacionado a la redacción de reportes breves de avance de proyecto.

Capítulo 7

Conclusiones

Los objetivos específicos de esta investigación se declaran logrados:

1. Se logró la recuperación de información cualitativa y cuantitativa, referente tanto al mercado laboral como al plan de formación interno en innovación de la carrera de Ingeniería Civil Eléctrica. Esto se ve reflejado en la primera etapa del proyecto (Recolección de datos).
2. Las etapas 2, 3 y 4 de este trabajo de título (Preparación de los datos, Etiquetado de la base de datos y Construcción de la propuesta, respectivamente), responden al segundo objetivo específico. Los resultados de estas etapas permiten afirmar que el objetivo fue logrado exitosamente.
3. La quinta etapa de este trabajo, Consolidación de la propuesta, así como el análisis realizado en el Capítulo 6 denotan el cumplimiento del tercer objetivo específico.

Para concluir esta investigación es necesario revisar cada una de las fases involucradas.

En la etapa de recolección de datos, es valioso destacar la posibilidad de recolección de información masiva del mercado laboral actual para futuros egresados de Ingeniería Eléctrica. Esta implementación, se puede expandir fácilmente a cualquier otra carrera, ya sea dentro de la Facultad de Ciencias Físicas y Matemáticas, como fuera de ella.

Se consideró un primer acercamiento a las herramientas de Inteligencia Artificial en la etapa del análisis del plan de formación actual de la carrera de Ingeniería Eléctrica, logrando una estructura de organización a partir de ellas. Este resultado es un buen punto de partida al buscar oportunidades de mejora para el plan de formación.

En la etapa de filtrado de la base de datos de ofertas de empleo, se realizó la comparación de rendimiento de dos distintas herramientas de Inteligencia Artificial en la tarea de clasificación de textos, logrando demostrar un destacable desempeño a favor de la Inteligencia Artificial de OpenAI.

A partir del análisis de las ofertas de empleo, se obtuvieron resultados contundentes respecto al mercado laboral actual. A partir de este, se destacan las habilidades asociadas al trabajo en equipo en proyectos tecnológicos, la identificación y análisis de problemas en proyectos tecnológicos y la confección de documentación técnica, como las más buscadas en el

mercado laboral.

Se utilizaron distintas herramientas de Inteligencia Artificial para generar agrupaciones de resultados de aprendizaje, con miras a la construcción de posibles cursos. Si bien existe una subsección dedicada a la consolidación final de la propuesta, cada parte del análisis de los resultados de agrupación es parte del entregable de esta investigación y aporta valor e información que antes no existía.

Se destacan los resultados obtenidos en la agrupación, ya que en varios de los resultados lograron agrupar según la estructura de organización creada inicialmente, a pesar de que no fue información ingresada a los modelos. Esto demuestra una convergencia en los modelos y posibles oportunidades de mejora.

La propuesta final puede resultar conservadora y, esto, se justifica en que fue realizada pensando en la posibilidad real de introducir alguna de las propuestas al plan de formación, siguiendo la estructura actual de los cursos enfocados a innovación de la carrera Ingeniería Civil Eléctrica de la Universidad de Chile. Sin embargo, se considera un aporte al desarrollo de nuestra carrera.

Mencionado todo lo anterior, se declara el objetivo general de la investigación como cumplido de forma exitosa.

7.1. Trabajo Futuro

Este trabajo entrega información relevante para el estudio del plan de formación en innovación de la carrera de Ingeniería Civil Eléctrica de la Universidad de Chile, abriendo la posibilidad a incluir las herramientas de Inteligencia Artificial dispuestas por distintas compañías en la planificación de cursos.

En el transcurso de la investigación se detectaron nuevas oportunidades de exploración en el área, las cuales quedan detalladas en los párrafos siguientes.

7.1.1. Inclusión de otros cursos del plan de formación

Para el desarrollo de esta investigación, se consideraron los 4 cursos referidos a la formación en innovación obligatorios del plan curricular de Ingeniería Civil Eléctrica. Quedaron fuera del alcance los siguientes grupos de cursos:

- Cursos de Plan Común.

Actualmente, existen cursos de plan común que forman la base de la formación en innovación para estudiantes de la carrera de Ingeniería Eléctrica. Estos cursos son: Desafíos de Innovación en Ingeniería y Ciencias (CD1100), Proyecto de Innovación en Ingeniería y Ciencias (CD1201) y Módulo Interdisciplinario (CD2201). Estos cursos podrían considerarse en el análisis ya que cubren parte de las temáticas que se deben revisar en la formación en innovación.

- Cursos electivos de Ingeniería Eléctrica.

Dentro los cursos disponibles como electivos en el catálogo de Ingeniería Eléctrica, se encuentran los cursos Seminario de Diseño e Innovación Tecnológica (EL4030, EL5030 y EL6030). Estos cursos están enfocados especialmente en el desarrollo de proyectos de innovación en el área de la Ingeniería Eléctrica, por lo que podrían ser incorporados en el análisis.

- Cursos del Magister de Innovación y Emprendimiento (MIE).

Los cursos del MIE son catalogados como electivos de especialidad en el plan de formación de Ingeniería Civil Eléctrica. Estos cursos ahondan en aspectos de la innovación y el emprendimiento con base científico-tecnológica, por lo que podrían ser incluidos en el análisis, especialmente pensando en una línea alternativa de cursos de innovación.

7.1.2. Identificar otras competencias a partir de herramientas de Inteligencia Artificial

A partir de las ofertas de empleo recuperadas para esta investigación, queda como propuesta ajustar los *prompt* para solicitar al modelo que identifique competencias que no están cubiertas por los resultados de aprendizaje pero que sí son solicitadas en el mercado laboral. Esta propuesta apunta a otro posible cierre de brechas entre lo buscado en el mercado laboral y lo considerado en el plan de formación de la carrera Ingeniería Civil Eléctrica.

En particular para esta propuesta, se realizaron algunas pruebas, identificando que el modelo es capaz de detectar nuevas habilidades, pero con resultados muy divergentes, por lo que es necesario iterar y ajustar el *prompt* para obtener resultados útiles.

7.1.3. Inclusión de opinión de estudiantes

A lo largo de este análisis, se consideró siempre que los resultados de aprendizaje de los programas de cada uno de los cursos son abordados exitosamente, pero en la práctica ¿están cubiertos efectivamente?.

A partir de la conversación con profesores del área, se obtuvo que es efectivo que dependiendo del desarrollo del semestre, no siempre son abordados todos los objetivos de los distintos ramos. Ante esta problemática, nace la posibilidad de incorporar dentro de los parámetros la opinión de los estudiantes de Ingeniería Eléctrica que toman los cursos y pueden proporcionar retroalimentación importante con el objetivo de lograr un cierre de brechas entre lo que es efectivamente abordado en los cursos y lo que se tiene como objetivo en los programas.

Utilizando herramientas de Inteligencia Artificial, se podrían detectar cuales resultados de aprendizaje podrían ser reubicados para obtener mejoras en la formación en innovación.

7.1.4. Ajuste de pesos por SCT

Hay muchos resultados de aprendizaje que, a pesar de que la redacción sea distinta, apuntan al mismo objetivo. Que los resultados de aprendizaje se repitan entre cursos no quiere decir que la relevancia o profundidad con que se aborda ese tema en particular, sea la misma para todos. Ante esto, resulta interesante incorporar la valoración de pesos que simbolice la dedicación que debería tener un resultado de aprendizaje en cada curso. Una buena referencia sería realizar esta escala a partir del Sistema de Créditos Transferibles (SCT).

Bibliografía

- [1] OECD, Learning for Jobs. OECD Reviews of Vocational Education and Training. OECD Publishing, 2010.
- [2] Holzer, H. J., “Skill mismatches in contemporary labor markets: How real? and what remedies?”, ILR Review, vol. 66, no. 3, pp. 503–531, 2013.
- [3] Conzelmann, J. G., Hemelt, S. W., Hershbein, B. J., Martin, S., y Simon, A., “Skills, majors, and jobs: Does higher education respond?”, University of North Carolina at Chapel Hill, 2024. See next page for additional authors.
- [4] López, H., Escalera, A., y García, C., “Personalización del aprendizaje con inteligencia artificial en la educación superior”, Revista Digital de Tecnologías Informáticas y Sistemas, vol. 7, pp. 123–128, 2023, doi:10.61530/redtis.vol7.n1.2023.165.123-128.
- [5] Berr, A., Leelaluk, S., Tang, C., Chen, L., Okubo, F., y Shimada, A., “Educational data analysis using generative ai”, 2024.
- [6] of Technology, M. I., “Reimagining and rethinking engineering education”. <https://news.mit.edu/2021/reimagining-rethinking-engineering-education-0715>, 2021.
- [7] of Engineering, N. A., “Innovations and opportunities in engineering education”. <https://www.nae.edu/113517/Innovations-and-Opportunities-in-Engineering-Education>, 2013.
- [8] Yang, H., Kim, J., y Lee, W., “Analyzing the alignment between ai curriculum and ai textbooks through text mining”, Applied Sciences, vol. 13, p. 10011, 2023, doi:10.3390/app131810011.
- [9] Vaitsis, C., Nilsson, G., y Zary, N., “Big data in medical informatics: Improving education through visual analytics.”, Studies in health technology and informatics, vol. 205, pp. 1163–1167, 2014, doi:10.3233/978-1-61499-432-9-1163.
- [10] Apablaza R., M., “Nuevo plan curricular de la fcfm: A la par de las instituciones internacionales”, Beauchef Magazine (Facultad de Ciencias Físicas y Matemáticas), vol. 12, no. 1, pp. 81–82, 2023.
- [11] Departamento de Ingeniería Eléctrica de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, “Perfil de egreso del ingeniero e ingeniera civil eléctrico (a)”, 2023, <https://www.die.cl/perfil-de-egreso-del-ingeniero-fcfm/>.
- [12] Departamento de Ingeniería Eléctrica, Universidad de Chile, EL3105 - Seminario de Ingeniería Eléctrica e Innovación Tecnológica., 2023. Programa del curso para [Primavera,2023].
- [13] Departamento de Ingeniería Eléctrica, Universidad de Chile, EL4204 - Formulación y Evaluación de Proyectos, 2023. Programa del curso para [Primavera,2023].

- [14] Departamento de Ingeniería Eléctrica, Universidad de Chile, EL6101 - Taller Proyectos Tecnológicos, 2023. Programa del curso para [Primavera,2023].
- [15] Departamento de Ingeniería Eléctrica, Universidad de Chile, EL6201 - Taller de Emprendimiento e Innovación, 2023. Programa del curso para [Otoño 2023].
- [16] Departamento de Ingeniería Eléctrica, Universidad de Chile, EL4030 - Seminario de Diseño e Innovación Tecnológica, 2011. Programa del curso para [Otoño 2011].
- [17] Departamento de Ingeniería Eléctrica, Universidad de Chile, EL5030 - Seminario de Diseño e Innovación Tecnológica II, 2011. Programa del curso para [Otoño 2011].
- [18] Departamento de Ingeniería Eléctrica, Universidad de Chile, EL6030 - Seminario de Diseño e Innovación Tecnológica III, 2011. Programa del curso para [Otoño 2011].
- [19] Matamala, M., Lasen Andrade, M., Chi-Duran, R., Caba Rutte, A., Patiño, M., Larroondo, J., y Meruane, V., “Beauchef proyecta: Implementación curricular de proyectos multidisciplinarios”, 2018.
- [20] Francisco Bugueño, “Metodología relámpago”, 2023. Gráfico utilizado con autorización del autor de la presentación de PowerPoint.
- [21] Osorio, C. A., “El arte de fallar”, Harvard Business Review, vol. 88, no. 4, pp. 76–85, 2010.
- [22] Bale, A. S., Ghorpade, N., S, R., Kamalesh, S., R, R., y S, R. B., “Web scraping approaches and their performance on modern websites”, en 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 956–959, 2022, doi:10.1109/ICESC54411.2022.9885689.
- [23] Felipe Jordão Almeida Prado Mattosinho, “Mining Product Opinions and Reviews on the Web”, 2010, https://www.rn.inf.tu-dresden.de/uploads/Studentische_Arbeiten/Masterarbeit_Mattosinho_Felipe.pdf.
- [24] SeleniumHQ, “Selenium - web browser automation”. <https://www.selenium.dev/>, 2024. Accessed: 2024-07-20.
- [25] Selenium, “The selenium browser automation project”, 2021, <https://www.selenium.dev/documentation/>.
- [26] IBM, “¿Qué es el procesamiento del lenguaje natural (NLP)?”, 2023, <https://www.ibm.com/es-es/topics/natural-language-processing>.
- [27] Jurafsky, D. y Martin, J., Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, vol. 2. 2008.
- [28] IBM, “Large language models”. <https://www.ibm.com/mx-es/topics/large-language-models>, 2024. Accessed: 2024-07-20.
- [29] Jurafsky, D. y Martin, J., Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition draft, vol. 3. 2024.
- [30] OpenAI, “Gpt-4 technical report”, 2024, <https://openai.com/research/gpt-4>.
- [31] Radford, A., Narasimhan, K., Salimans, T., y Sutskever, I., “Improving language understanding by generative pre-training”, OpenAI, 2018, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

- [32] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Nee-lakantan, A., Shyam, P., Sastry, G., Askell, A., *et al.*, “Language models are few-shot learners”, arXiv, 2020.
- [33] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., y Zettlemoyer, L., “Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension”, arXiv preprint arXiv:1910.13461, 2020.
- [34] Hugging Face, “Zero-shot classification”, <https://huggingface.co/tasks/zero-shot-classification>.
- [35] Face, H., “Tokenizers — transformers 4.6.1 documentation”. https://huggingface.co/docs/transformers/main_classes/tokenizer, 2024. Accessed: 2024-07-20.
- [36] SpaCy, “Library Architecture”, <https://spacy.io/api>.
- [37] Jolliffe, I. T., Principal Component Analysis, vol. 2. 2002.
- [38] IBM, “What is unsupervised learning?”. <https://www.ibm.com/topics/unsupervised-learning>, 2024. Accessed: 2024-07-20.
- [39] Manning, C. D., Raghavan, P., y Schütze, H., Introduction to Information Retrieval. Cambridge University Press, 2008.
- [40] Han, J., Kamber, M., y Pei, J., Data Mining: Concepts and Techniques. Elsevier Inc, 2012.

Anexo A

Resultados

A.1. Descripción de Resultados de Aprendizaje

Tabla A.1: Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.

N	Resultado de Aprendizaje	Etiqueta
1	Analiza un desafío propuesto y el contexto en que se inserta (cultural, social ambiental y económico), considerando las etapas del proceso de innovación para proponer ideas y seleccionar aquella que pueda transformarse en un potencial proyecto plasmable en un prototipo de baja resolución.	Análisis de Desafío Innovación
2	Define el problema, a partir del desafío propuesto, con su correspondiente justificación, gestionando la información del usuario, entorno y ecosistema, a través de un análisis de lo existente de la industria, a fin de identificar potenciales oportunidades del proyecto y su impacto.	Identificación de Problemas y Oportunidades
3	Diseña e implementa, con criterio de innovación, un prototipo de baja resolución que demuestre total o parcialmente la solución de carácter tecnológico para un problema y su potencial impacto, plasmable en diferentes formatos como software, hardware, modelos o cualquier soporte audiovisual.	Diseño e Implementación de Prototipos Innovadores
4	Formula proyectos de inversión públicos y privados en ingeniería eléctrica, considerando etapas y componentes e integrando a su formulación criterios técnicos, económicos, sociales, ambientales y éticos.	Formulación de Proyectos de Inversión
5	Calcula indicadores de rentabilidad de un proyecto en ingeniería eléctrica, desde una perspectiva privada y social, a fin de determinar su viabilidad.	Cálculo de Rentabilidad de Proyectos
6	Evalúa proyectos de inversión públicos y privados en ingeniería eléctrica a partir del análisis de criterios e indicadores respecto de sus beneficios y costos en materias técnicas, económicas, ambientales, sociales y éticas.	Evaluación de Proyectos de Inversión
7	Elabora un proyecto en el ámbito de la Ingeniería eléctrica, conforme a los requerimientos correspondientes a cada una de las etapas de este, considerando especificaciones técnicas, normativas, protocolos, así como requerimientos económicos, ambientales, sociales y éticos.	Elaboración de Proyectos de Ingeniería

(Continúa en la página siguiente)

Tabla A.1: Continuación. Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.

N	Resultado de Aprendizaje	Etiqueta
8	Evalúa tiempos, costos de procesos productivos de piezas y componentes eléctricos, considerando tecnologías y procesos de manufactura disponibles, diseño de prototipos, experimentos e información de mercado, etc., para proponer opciones de mejora.	Evaluación de Costos y Tiempos de Producción
9	Evalúa la factibilidad económica de cada etapa del estudio de ingeniería y/o el desarrollo de un producto, considerando las condiciones de mercado y licitaciones posibles y sugiriendo mejoras de ser necesario.	Factibilidad Económica de Proyectos
10	Elabora documentación técnica, planos y memorias de cálculo, considerando rigurosidad en el manejo de estándares y buenas prácticas asociadas a la ética profesional en ingeniería, los que integra a la propuesta de proyecto en cada una de las áreas de especialidad.	Elaboración de Documentación Técnica
11	Diseña y proyecta una solución a un problema, utilizando metodologías ágiles para productos, ingeniería de proyecto de innovación y de integración, y para proyectos tradicionales, a fin de agregar valor a un sistema o proceso que permita mejoras al bienestar humano.	Metodologías Ágiles en Diseño de Soluciones
12	Define herramientas de emprendimiento con las cuales trabajará el dolor o necesidad, por ejemplo, Design Thinking, maquetas (mock up) y canvas, entre otros.	Uso de Herramientas de Emprendimiento
13	Evalúa el potencial de la oportunidad aprovechable, a partir de la experiencia realizada y los mock up o prototipos que haya presentado, en términos de su valor social y económico y de su eficiencia, considerando a los usuarios y/o clientes.	Evaluación de Oportunidades y Prototipos
14	Analiza problemas operativos de un proyecto, desde la perspectiva de gestión y emprendimiento y que deben ser abordados desde lo ingenieril.	Análisis de Problemas Operativos
15	Valida comercialmente, en una primera iteración, el dolor, mediante un mock up, prototipado, simulación de ventas, entre otros.	Validación Comercial de Prototipos
16	Levanta evidencia que sustente la estimación de mercado, a partir de las definiciones y técnicas tales como TAM, SAM y SOM, considerando, por ejemplo, espacio geográfico, tendencias, barreras de entrada, entre otros.	Estimación de Mercado y Barreras
17	Construye una primera versión de plan de negocios, mediante herramientas como, por ejemplo, Design Thinking, maquetas (mock up), Lean Launchpad, metodologías de Business Model Canvas o Lean Canvas, Modelo de desarrollo de clientes y metodologías ágiles, considerando los recursos clave y su disponibilidad.	Construcción de Plan de Negocios
18	Diseña un modelo de negocios para introducir el producto mínimo viable (PMV), con valor social o económico, validando las hipótesis.	Diseño de Modelos de Negocio

(Continúa en la página siguiente)

Tabla A.1: Continuación. Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.

N	Resultado de Aprendizaje	Etiqueta
19	Identifica un dolor o necesidad de la sociedad y de las personas sobre el cual se trabajará, a través de una prospección del problema, considerando el mercado actual y sus variaciones futuras, opciones de financiamiento, impacto social, cultural y económico.	Prospección de Necesidades Sociales
20	Evalúa si el proyecto posee tracción, es decir, si se verifica un interés real de los potenciales usuarios o clientes.	Validación de Tracción de Proyecto
21	Trabaja en equipo en un proyecto tecnológico, donde se definen roles, se planifica y organiza cada actividad con su entregables comprometido, estableciendo una comunicación efectiva, manejo de conflicto para concordar acuerdos que favorezcan la toma de decisiones.	Trabajo en Equipo en Proyectos Tecnológicos
22	Redacta, de manera clara y eficaz, informes de avance y final del proyecto, considerando dominio conceptual y una fundamentación rigurosa sobre las decisiones de diseño, planos, especificaciones técnicas, desarrollo de prototipos, procesos de manufactura, selección óptima de materiales y cumplimiento de normativas vigentes.	Redacción de Informes de Proyecto
23	Expone, en forma oral, sobre la solución dada a un problema de ingeniería, a partir de argumentos técnico-disciplinarios que sustenten la propuesta, evidenciando en su discurso claridad, uso de un lenguaje variado y precisión conceptual.	Presentación de Soluciones Técnicas
24	Crea, con sus pares, el equipo de trabajo, concordando roles y funciones, definir objetivos comunes, establecer acciones, plazos y recursos.	Creación y Gestión de Equipos
25	Comprende y orienta su trabajo hacia el propósito colectivo, comprendiendo que este está por sobre la agenda personal	Orientación al Propósito Colectivo
26	Presenta en forma oral y escrita aspectos y conceptos centrales de su proyecto, considerando en su exposición una comunicación efectiva y empática que conecte con el tipo de audiencia	Comunicación Efectiva de Proyectos
27	Reevalúa la propuesta de valor, en base a las pruebas e hipótesis que ha ido planteando, desarrollando una comprensión de la necesidad o dolor	Reevaluación de Propuesta de Valor
28	Propone distintas soluciones, las que va testeando a través de maquetas	Testeo de Soluciones con Maquetas
29	Evalúa, de manera crítica, las acciones propias y las del equipo, y cómo estas repercuten en el desarrollo del proyecto	Evaluación Crítica de Acciones

(Continúa en la página siguiente)

Tabla A.1: Continuación. Resultados de aprendizaje relacionados a su etiqueta e índice correspondiente.

N	Resultado de Aprendizaje	Etiqueta
30	Identifica, a través de una introspección, problemas que surgen en el trabajo con los miembros de su equipo, así como las aptitudes de cada uno, proponiendo soluciones	Identificación y Solución de Problemas de Equipo
31	Evalúa, desde lo económico y social, el producto	Evaluación Económica y Social de Productos
32	Valida el mínimo producto viable y su demanda, considerando prueba y error, mercado, definición y dinámica del producto o servicio e identificando puntos de mejora y posibles inconvenientes que puedan surgir.	Validación de Producto Mínimo Viable
33	Presenta su propuesta, mediante pitch y exposiciones orales, considerando el tipo de audiencia a la que se enfrenta, precisión y fluidez en el desarrollo de sus ideas, así como capacidad de persuasión y síntesis	Pitch y Exposiciones Orales
34	Expone en forma oral y escrita avances y resultados de la formulación y evaluación de un proyecto de inversión en ingeniería eléctrica, informando, de manera clara y coherente, el procedimiento ejecutado y las conclusiones derivadas de dicha formulación y evaluación.	Exposición de Proyectos de Inversión
35	Produce, en forma escrita reportes de avance e informe técnico del prototipo de baja resolución con su correspondiente proceso de diseño, fundamentando sus decisiones con argumentos que respalden el análisis de antecedentes, las etapas de diseño y ventajas competitivas del prototipo.	Producción de Reportes Técnicos
36	Expone, con apoyo de recursos audiovisuales, los resultados de su proyecto de innovación, combinando con destreza, recursos no verbales (expresión corporal, mirada, gestos) con recursos verbales (lenguaje claro, variado y fluido), para convencer sobre el valor de la propuesta.	Presentación con Recursos Audiovisuales
37	Evalúa la propuesta de solución al desafío planteado, considerando metodologías de innovación, así como un análisis de contexto en que este se inserta, y aspectos técnicos, económicos y de factibilidad, entre otros.	Evaluación Integral de Soluciones

(Fin de la tabla)

A.2. Agrupación de Resultados de Aprendizaje

A.2.1. Agrupación con GPT-4o (Respuesta)

A continuación, se adjunta la respuesta completa dada por el modelo GPT-4o ante la solicitud de agrupar los resultados de aprendizaje para formar cursos que puedan ser incorporados al plan de formación de la carrera de Ingeniería Civil Eléctrica. Es importante mencionar que no fue cambiado ningún carácter de la respuesta entregada por el modelo.

Para construir una propuesta de planes de formación organizados en cursos que aborden la temática de innovación dentro de la carrera de Ingeniería Civil Eléctrica, es importante tomar en cuenta las competencias generales, subcompetencias y los pesos escalados. La si-

guiente es una propuesta de agrupación en base a la información proporcionada:

Curso 1: **Análisis e Identificación de Problemas y Oportunidades Innovadoras**

En este curso se trabaja en la identificación y análisis de problemas y oportunidades, con énfasis en entender y contextualizar los desafíos. También incluye competencias de comunicación y trabajo en equipo integradas en los proyectos.

- Resultados de aprendizaje:
- 0. Analiza un desafío propuesto y el contexto en el que se desarrolla.
- 1. Define el problema, a partir del desafío propuesto.
- 18. Identifica un dolor o necesidad de la sociedad.
- 26. Reevalúa la propuesta de valor, en base a las observaciones de los test.
- 6. Elabora un proyecto en el ámbito de la Ingeniería Civil Eléctrica.
- 21. Redacta, de manera clara y eficaz, informes de proyecto.
- 22. Expone, en forma oral, sobre la solución dada al problema.
- 23. Crea, con sus pares, el equipo de trabajo, con respeto e inclusión de sus miembros.

Curso 2: **Diseño y Prototipado de Soluciones Innovadoras**

Este curso se enfoca en diseñar y desarrollar prototipos y evaluar sus funcionalidades utilizando metodologías ágiles.

- Resultados de aprendizaje:
- 2. Diseña e implementa, con criterio de innovación, soluciones tecnológicas ante un problema identificado.
- 3. Formula proyectos de inversión públicos y privados.
- 14. Valida comercialmente, en una primera iteración, el prototipo de la solución propuesta.
- 27. Propone distintas soluciones, las que va testeando en función de los requerimientos descritos en cada caso.
- 31. Valida el mínimo producto viable y su demanda, según criterios del desafío planteado.
- 10. Diseña y proyecta una solución a un problema, usando metodologías ágiles.
- 33. Expone en forma oral y escrita avances y resultados de un proyecto de inversión.
- 7. Evalúa tiempos, costos de procesos productivos y su implementación.
- 28. Evalúa, de manera crítica, las acciones propias y de sus pares.

Curso 3: **Evaluación y Gestión de Proyectos Innovadores**

Aquí se abordan las herramientas y metodologías para la gestión y evaluación de proyectos con criterios de eficiencia y sostenibilidad.

- Resultados de aprendizaje:
- 4. Calcula indicadores de rentabilidad de un proyecto de inversión en ingeniería.
- 5. Evalúa proyectos de inversión públicos y privados usando técnicas avanzadas.
- 8. Evalúa la factibilidad económica de cada etapa de un proyecto de ingeniería.
- 30. Evalúa, desde lo económico y social, el producto.

- 32. Presenta su propuesta, mediante pitch y exposición oral eficaz.
- 12. Evalúa el potencial de la oportunidad aprovechando la herramienta del prototipo.
- 17. Diseña un modelo de negocios para introducir el proyecto en el mercado.
- 15. Levanta evidencia que sustente la estimación de las barreras del mercado.

Curso 4: ****Implementación y Validación de Proyectos****

En este curso se enfoca en la implementación, monitoreo y validación de proyectos, evaluando sus impactos y realizando iteraciones.

- Resultados de aprendizaje:
 - 11. Define herramientas de emprendimiento con las cuales podrá aprovechar oportunidades y crear iniciativas sustentables.
 - 13. Analiza problemas operativos de un proyecto, determinando sus causas fundamentales.
 - 16. Construye una primera versión de plan de negocio, considerando la rentabilidad del producto.
 - 9. Elabora documentación técnica, planos y memoria descriptiva del proyecto de ingeniería.
 - 20. Trabaja en equipo en un proyecto tecnológico, demostrando compromiso y colaboración.
 - 25. Presenta en forma oral y escrita aspectos y conclusiones del avance del proyecto.
 - 35. Expone, con apoyo de recursos audiovisuales, los resultados y fases del trabajo.
 - 34. Produce, en forma escrita reportes de avance en el proyecto.

Curso 5: ****Monitoreo y Evaluación Continua de Proyectos****

Enfocado en la evaluación continua y monitoreo para asegurar el éxito del proyecto hasta su fase de consolidación y posterior mejora.

- Resultados de aprendizaje:
 - 19. Evalúa si el proyecto posee tracción, es decir tiene utilidad real.
 - 24. Comprende y orienta su trabajo hacia el propósito colectivo.
 - 29. Identifica, a través de una introspección, problemas operacionales y de equipo, y propone soluciones.
 - 36. Evalúa la propuesta de solución al desafío planteado en los diferentes procesos del ciclo del proyecto.
 - 37. Propone acciones de mejora continua basándose en la evaluación integral del proyecto y su sostenibilidad.

—
****Justificación de eliminación o repetición:****

No se elimina ningún resultado de aprendizaje, pero algunos se repiten en diferentes cursos para asegurar una formación continua y consistente en las competencias transversales de comunicación y trabajo en equipo, como se especificó.

****Distribución del peso escalado:****

Esta distribución busca balancear los pesos de los resultados de aprendizaje, garantizando que en cada curso se traten aspectos esenciales demandados por el mercado laboral, como la rentabilidad, la factibilidad económica y la evaluación crítica de proyectos.

Cada curso integra módulos de comunicación y trabajo en equipo para asegurar el desarrollo progresivo de estas competencias en cada etapa formativa.