



ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS.

**INNOVACIÓN Y SINERGIA PARA LA EDUCACIÓN SUPERIOR
ECUATORIANA: DESARROLLO DE UN MARCO DE INTELIGENCIA DE
NEGOCIOS PARA LA OPTIMIZACIÓN DE ALIANZAS EMPRESARIALES Y
ORIENTACIÓN ESTUDIANTIL PARA EL AÑO 2023**

Profesor

Victor Manuel Gonzalez Holguin

Autor

Carlos Magno Brusil Gavela

2023

RESUMEN

El presente trabajo se centra en abordar la actual brecha que existe entre la formación académica y la empleabilidad de los graduados en Ecuador, destacando la necesidad de integrar experiencias laborales prácticas y habilidades demandadas por el mercado en los currículos de los universitarios. Para cerrar esta brecha se busca utilizar metodologías de inteligencia de negocios para evaluar y mejorar las relaciones corporativas y el desarrollo profesional de los estudiantes. Se implementan modelos de clustering para optimizar la gestión de relaciones con empresas y un sistema de recomendación para alinear las competencias de los estudiantes con las necesidades del mercado. Los resultados sugieren que la adaptación de estrategias educativas y la colaboración con el sector empresarial son cruciales para mejorar la empleabilidad de los graduados.

ABSTRACT

The current project focuses on addressing the existing gap between academic training and the employability of graduates in Ecuador, highlighting the need to integrate practical work experiences and skills demanded by the market into the university curricula. To close this gap, it seeks to use business intelligence methodologies to evaluate and improve corporate relationships and the professional development of students. Clustering models are implemented to optimize the management of relationships with companies and a recommendation system to align student's competencies with market needs. The results suggest that the adaptation of educational strategies and collaboration with the real sector are crucial for improving the employability of graduates.

Índice de Contenidos

| | |
|---|----|
| RESUMEN | 2 |
| ABSTRACT | 3 |
| 2. Revisión de Literatura | 6 |
| 3. Objetivo del estudio | 10 |
| 4. Definición del problema | 12 |
| 5. Objetivo Principal | 13 |
| 6. Objetivos Específicos | 14 |
| 7. Justificación y Aplicación de la Metodología | 14 |
| 7.1. Selección de la base de datos | 15 |
| 7.1.1. Bases de datos para la creación de relaciones corporativas | 16 |
| 7.1.2. Bases de datos para la aplicación de microaprendizaje en el desarrollo profesional de estudiante | 16 |
| 7.2 Limpieza, preprocesamiento y transformación de datos | 17 |
| 7.2.1 Preprocesamiento y transformación de las bases de datos para la creación de relaciones corporativas | 19 |
| 7.2.2 Preprocesamiento y transformación de datos para la aplicación de microaprendizaje en el desarrollo profesional de estudiante | 22 |
| 7.3. Identificación y descripción de variables | 24 |
| 7.4. Selección de los modelos | 35 |
| 7.4.1 Clustering | 36 |
| 7.4.2 Sistema de recomendación | 39 |
| 8. Resultados | 41 |
| 9. Estrategia Organizacional | 47 |
| Anexo | 59 |

Índice de Figuras

| | |
|---|----|
| Figura 1: Número de ofertas laborales durante el periodo 2022-2023 | 25 |
| Figura 2: Histograma del número de ofertas laborales por categoría..... | 26 |
| Figura 3: Nube de palabras basado en la descripción del empleo. | 27 |
| Figura 4: Distribución de las empresas por sector CIIU nivel 1 | 28 |
| Figura 5: Distribución por tamaño de la empresa..... | 29 |
| Figura 6: Mapa de calor sobre la cantidad de solicitudes por empleador y tipo de empleo..... | 30 |
| Figura 7: Clasificación de los cursos por su área de conocimiento, enrollment y calificación..... | 31 |
| Figura 8: Grafico de bigotes | 32 |
| Figura 9: Grafico de bigotes | 32 |
| Figura 10: Grafico de bigotes | 33 |
| Figura 11: Transformación logarítmica de la variable empleados | 34 |
| Figura 12: Transformación logarítmica del impuesto a la renta causado | 34 |

Índice de Tablas

| | |
|--|----|
| Tabla 1: Base de datos del Carrer Services del directorio de empresas registrado en la plataforma..... | 19 |
| Tabla 2: Base de datos del Ranking Empresarial 2023..... | 20 |
| Tabla 3: Base final después de homologación | 21 |
| Tabla 4: Base de datos de plataforma de los registros de las postulaciones anuales 2022-2023..... | 22 |
| Tabla 5: Base datos de cursos UDEMY 2023 | 23 |
| Tabla 6: Base del Carrer Service transformada | 24 |
| Tabla 7: Resultado de los resultados Silhouette Score | 43 |
| Tabla 8: Resultado de los resultados Calinski-Harabasz | 45 |

1.Introducción

La educación superior en Ecuador enfrenta el desafío de adaptarse a las nuevas demandas de la empleabilidad de sus graduados, un problema exacerbado por la rápida evolución del mercado laboral y la creciente demanda de habilidades que trascienden el conocimiento técnico. La revisión de la literatura subraya la importancia de la empleabilidad percibida y la necesidad de integrar habilidades prácticas y teóricas en la formación académica, destacando la adopción de metodologías innovadoras como el microaprendizaje y la enseñanza invertida.

El objetivo de este estudio se centra en buscar opciones para minimizar las brechas entre la formación académica universitaria tradicional y la empleabilidad de los graduados, enfatizando la necesidad de integrar prácticas laborales y habilidades tecnológicas y blandas en los currículos. La adaptabilidad de los currículos y la innovación en los métodos de enseñanza son cruciales para aumentar la empleabilidad y adaptabilidad de los graduados a entornos laborales volátiles.

La definición del problema señala una creciente discrepancia entre las habilidades obtenidas por los graduados y las necesidades del mercado laboral, lo que pone en riesgo la efectividad de la educación superior. Este problema multifacético requiere un enfoque integral que combine el desarrollo de competencias técnicas y habilidades blandas con estrategias educativas innovadoras y una estrecha colaboración con el sector empresarial.

En cuanto a los modelos de análisis de datos, se utilizan dos modelos de clustering y un sistema de recomendación para alinear las habilidades de los estudiantes con las demandas del mercado laboral. Los modelos de clustering ayudan a optimizar la gestión de relaciones con empresas, mientras que el sistema de recomendación personaliza la orientación profesional para los estudiantes. Estos modelos reflejan un compromiso con la aplicación práctica y orientada al análisis de datos para mejorar la inserción profesional de los graduados.

2. Revisión de Literatura

El sistema de educación superior ecuatoriano, como en muchas partes del mundo, enfrenta el desafío de mejorar la empleabilidad de sus graduados. Este análisis busca presentar estrategias basadas en la literatura académica reciente para mejorar la inserción laboral de la comunidad universitaria. (Hu, Jiang, & Chen, 2022; Lázár et al., 2023; Ressia & Shaw, 2022). La investigación de Hu, Jiang y Chen (2022) muestra la importancia de la empleabilidad percibida cuando existe seguridad laboral, sugiriendo que cuando se trabaja en mejorar la percepción de empleabilidad a través de relaciones corporativas efectivas la empleabilidad de los estudiante universitarios y graduados. Complementando estas ideas podemos encontrar autores como, Lázár et al. (2023) que revisa y analiza las competencias de comunicación intercultural, identificando ventajas y desventajas para los empleadores, en el estudio se resalta la necesidad de habilidades que trasciendan el conocimiento técnico en un mercado que cada vez tiene menos fronteras. (Hu, Jiang, & Chen, 2022; Lázár et al., 2023; Cadet & Griffiths, 2023).

La preparación en temas de empleabilidad para la inserción laboral, según Ressia y Shaw (2022), muestra una necesidad de integrar no solo elementos teóricos, sino también elementos prácticos a la formación de los futuros profesionales (Ressia & Shaw, 2022; Flornoy-Guédon et al., 2023; Blinov, Sergeev, & Rodichev, 2022), dentro de su investigación sugiere la adopción de metodologías innovadoras como lo es el microaprendizaje. Es para este apartado donde el trabajo de Flornoy-Guédon et al. (2023) sobre el microaprendizaje en la formación continua apoyaría su aplicación en el desarrollo profesional de los estudiantes ecuatorianos. Fidan (2023) se suma en la demostración de la efectividad de utilizar metodologías de aula invertida apoyada por microaprendizaje, lo que sugiere su potencial para mejorar la preparación de los estudiantes para el mercado laboral. (Flornoy-Guédon et al., 2023; Hanshaw & Hanson, 2019; Fidan, 2023).

Mientras tanto Draissi et al. (2023) y Abelha et al. (2020) señalan la existencia de una brecha significativa entre las competencias con las que los graduados salen de su formación académica superior y las necesidades del mercado laboral, lo que subraya la importancia de alinear los currículos y estrategias de desarrollo profesional con las demandas del mercado actual. (Fidan, 2023; Draissi et al., 2023; Abelha et al., 2020). Este desajuste presenta una necesidad esencial para el uso de herramientas como el Balanced Scorecard como potenciador de las áreas de relaciones corporativas y desarrollo profesional. Hayden y Ledwith (2014) proporcionan un marco para el fortalecimiento de relaciones corporativas para el sector de la educación superior, destacando la importancia de los servicios de carrera en las relaciones externas universitarias (Draissi et al., 2023; Gooptu, Bros, & Chowdhury, 2023; Elkosantini et al., 2023).

Combinar estrategias de relaciones externas y desarrollo profesional es imperativo para abordar el desafío de la empleabilidad de los jóvenes en Ecuador. El microaprendizaje, como herramienta según Blinov, Sergeev y Rodichev (2022), crea una oportunidad para personalizar la preparación de los estudiantes según las necesidades de diferentes empresas. Además, el fortalecimiento de las relaciones corporativas eficientes mejora la comprensión de las demandas del mercado y orienta los programas de desarrollo profesional hacia las necesidades deseadas (Hayden & Ledwith, 2014; Manlagaylay, 2022; Buckholtz & Donald, 2022).

Innovación es una de las palabras en las que se basó la revisión de la literatura revisada, con este antecedente surgieron nuevas estrategias como el microaprendizaje y el aula invertida, temas que son relevantes cuando se habla de la preparación de los estudiantes para los retos de un mercado laboral en constante cambio. La evaluación continua de estas estrategias garantizará su efectividad y relevancia en el cambiante panorama del mercado laboral. Un ejemplo del uso de estas estrategias es la investigación de Cadet y Griffiths (2023) demuestra cómo la integración de la empleabilidad en el currículo de Ciencias Sociales puede mejorar la preparación de los estudiantes. Además, Sachitra y Kumara (2019) encontraron

que los graduados en programas de gestión tenían una alta tasa de empleabilidad, por el tipo de currículum que maneja la carrera sugiere la importancia de desarrollar capacidades personales además de logros académicos (Fidan, 2023; Cadet & Griffiths, 2023; Sachitra & Kumara, 2019).

Otro estudio importante es el de Manlagaylay (2022) sobre la implementación efectiva de servicios de desarrollo de carrera dentro de las instituciones educativas y colocación laboral, el estudio el sujeto de estudio escogido en el estudio fue una universidad estatal, donde se presenta indicios de la importancia de la sostenibilidad de implementar estos servicios en diversas plataformas y formatos. Buckholtz y Donald (2022) resaltan la teoría de los resultados directos que se obtiene al incorporar a actores como el mediador en las relaciones entre asesores de carrera universitarios y reclutadores de graduados, lo que permite una perspectiva innovadora sobre cómo mejorar la colaboración y comunicación entre estas partes obteniendo un resultado 360 del mercado laboral (Manlagaylay, 2022; Buckholtz & Donald, 2022).

Para enfrentar el desafío de la empleabilidad, las universidades ecuatorianas, las organizaciones del sector de enseñanza superior deben adoptar un enfoque integral que combine el desarrollo de competencias técnicas y habilidades blandas. Este enfoque debe ir acompañado de estrategias educativas innovadoras y una estrecha colaboración con el sector empresarial. Las metodologías como el microaprendizaje y el aula invertida, así como un enfoque en habilidades de comunicación intercultural y desarrollo personal, son fundamentales (Blinov, Sergeev, & Rodichev, 2022; Fidan, 2023; Lázár et al., 2023).

En conclusión, este análisis integrado hace hincapié en la necesidad de una estrategia coherente en la educación superior que combine eficazmente las relaciones corporativas con programas de desarrollo profesional orientados a mejorar la empleabilidad de los graduados. La adaptación continua de estas estrategias a las necesidades cambiantes del mercado laboral es crucial. Las

universidades deben ser proactivas en la formación de profesionales que no solo sean competentes en su campo de estudio, sino que también estén equipados con las habilidades y competencias demandadas en el entorno laboral actual. La colaboración entre universidades y el sector empresarial, junto con un enfoque en metodologías educativas innovadoras, jugará un papel crucial en la preparación de los estudiantes para los retos del mercado laboral en Ecuador.

3. Objetivo del estudio

El objeto de estudio de este proyecto se centra en la diferencia que existe entre la formación académica universitaria y la empleabilidad de los graduados en la educación superior, un desafío que ha cobrado mayor relevancia en el contexto actual, tomando en cuenta el desarrollo de nuevas tecnologías que han acelerado la globalización. El estudio de Portillo-Navarro, Lagos-Rodríguez y Meseguer-Santamaría (2022) resalta la importancia del desarrollo de experiencia laboral durante los estudios, demuestran una mejora en la empleabilidad y la calidad del empleo entre los graduados. Este enfoque enfatiza la necesidad de integrar prácticas laborales y nuevas habilidades demandadas como, lo son el uso de tecnologías en los currículos universitarios, lo que es crucial para aumentar las oportunidades laborales de los graduados (Portillo-Navarro, Lagos-Rodríguez, & Meseguer-Santamaría, 2022).

Durante la revisión narrativa de modelos conceptuales de empleabilidad realizada por Cook (2022) permite observar un marco para comprender la complejidad de la empleabilidad en la educación superior. Este análisis subraya que existen variables como la percepción sobre carreras, empleabilidad y empleo, las cuales tienen una variabilidad considerable, que se ve influenciada directamente por la enseñanza y los currículos dictados por las instituciones. La adopción de los modelos conceptuales revisados durante el texto se explica que las estrategias educativas

pueden proporcionar una base más sólida para diseñar programas afecten de manera positiva a la empleabilidad de la comunidad universitaria (Cook, 2022).

La adaptabilidad de los currículos de los programas universitarios para incluir experiencias laborales prácticas es una iniciativa que afecta de manera directa a los hallazgos de Portillo-Navarro et al. (2022). Al incorporar estas experiencias y habilidades que se encuentran en la demanda de las empresas en la formación de los estudiantes, las instituciones de educación superior pueden preparar mejor a los graduados para los requisitos y desafíos del mercado laboral actual. Además, este enfoque puede ayudar a reducir la brecha de empleabilidad (Portillo-Navarro et al., 2022).

La importancia de la innovación en los métodos y programas de enseñanza tienen un rol importante para el contexto. Siguiendo el punto de vista de Cook (2022), se insta a las universidades a promover metodologías de enseñanza avanzadas, como el microaprendizaje y el aula invertida, que son efectivas para mejorar tanto la relevancia como la aplicabilidad de la educación superior. Estas metodologías no solo aumentan la empleabilidad, sino que también mejoran su capacidad para adaptarse a entornos laborales volátiles (Cook, 2022).

La importancia de este proyecto es subrayar la necesidad de evaluación y adaptación continua de las estrategias corporativas y la adaptabilidad de los programas educativos. Siguiendo los hallazgos empíricos, las universidades deben establecer sistemas de retroalimentación y evaluación que aseguren el éxito de sus estrategias de desarrollo profesional y las estrategias de relaciones con el sector empresarial sean efectivas y mantengan una concordancia con las nuevas demandas del mercado laboral. Este enfoque holístico es el corazón del proyecto que busca cerrar la brecha entre los sistemas convencionales de educación y la empleabilidad en el mundo contemporáneo (Portillo-Navarro et al., 2022; Cook, 2022).

4. Definición del problema

La actualidad de la educación superior presenta una creciente discrepancia entre las habilidades adquiridas por los graduados y las demandas del mercado laboral. Esta brecha representa un problema multifacético que tiene repercusiones tanto a las instituciones educativas como a los estudiantes universitarios, poniendo en riesgo la relevancia y la efectividad de los métodos de enseñanza en el panorama educativo contemporáneo (Draissi et al., 2023; Abelha et al., 2020). La naturaleza de este problema es compleja, involucrando la vertiginosa evolución de las necesidades del mercado y la obsolescencia acelerada de los enfoques educativos utilizados en las diferentes instituciones. Las diferencias de enfoque y necesidades la educación y las expectativas prácticas del entorno laboral exige una respuesta educativa eficiente, que permite generar una verdadera sinergia (Hayden & Ledwith, 2014).

Este problema es particularmente importante para las instituciones de educación superior, la razón del ¿Por qué la empleabilidad es crucial para las universidades?, se ve relacionado directamente con la relevancia y efectividad que pueden mantener en un mercado laboral en constante transformación. La disonancia en habilidades tiene implicaciones significativas en la empleabilidad de los graduados, lo que puede llevar a una desvalorización del valor percibido de la educación superior (Fidan, 2023). Por ello, es fundamental adoptar un enfoque analítico que permita una comprensión profunda de esta brecha y facilite el desarrollo de estrategias informadas para abordarla de manera efectiva (Wulk, s.f.).

Uno de los principales ejes de este proyecto es desarrollar e implementar estrategias integradas en las relaciones externas y programas de desarrollo profesional en universidades, para mejorar la empleabilidad de los graduados. Este objetivo se centra en ir acortando la brecha entre la formación académica convencional y las demandas del mercado laboral, asegurando que los

estudiantes adquieran habilidades relevantes y actualizadas que aumenten sus posibilidades de obtener la plaza de empleo deseada tras graduarse (Kaminski et al., s.f.). Para tener éxito en esta meta es esencial poder tener una visión multidimensional como lo es la evaluación la efectividad de los programas actuales, fortalecer las relaciones con el sector empresarial, implementar metodologías educativas innovadoras como el microaprendizaje y aulas invertidas, desarrollar programas de capacitación personalizados, realizar evaluaciones continuas y promover la adaptabilidad y flexibilidad en la enseñanza tanto de los docentes como los estudiantes. (Hanshaw & Hanson, 2019).

En conclusión, la efectividad de este enfoque integrado no solo beneficiará a los estudiantes y graduados en términos de empleabilidad, sino que también permitirá potenciar la posición de las universidades como instituciones clave en la preparación de individuos más aptos para el mercado laboral contemporáneo. La realización de estos objetivos requiere un compromiso con la innovación y la colaboración, asegurando que las estrategias implementadas sean relevantes y efectivas frente a las tendencias cambiantes del mercado laboral (Senarath & Patabendige, 2015).

5. Objetivo Principal

Mejorar de manera integral la empleabilidad de los graduados universitarios mediante el desarrollo y la implementación de estrategias integradas en relaciones externas y programas de desarrollo profesional. Este objetivo busca cerrar efectivamente la brecha entre las habilidades y competencias adquiridas por los estudiantes y las demandas del mercado laboral contemporáneo, asegurando así que los graduados estén mejor preparados para enfrentar los desafíos y aprovechar las oportunidades en sus respectivos campos profesionales.

6. Objetivos Específicos

1. **Aplicación de Modelos de Clustering en la Gestión de Relaciones Estratégicas:** Implementar y comparar dos modelos de clustering distintos para analizar y mejorar la gestión de las relaciones con empresas y organizaciones relevantes. Este enfoque tiene como fin identificar el modelo más efectivo para desarrollar alianzas estratégicas que respondan a las necesidades actuales del mercado laboral, facilitando así la inserción profesional de los graduados.
2. **Optimización de Estrategias de Colaboración mediante Criterios de Clustering:** Utilizar los modelos de clustering para establecer criterios que permitan optimizar la selección y mantenimiento de alianzas estratégicas con el sector empresarial. Estos criterios se enfocarán a garantizar una alineación continua entre las competencias de los graduados y las expectativas del mercado, mejorando las oportunidades de prácticas y empleabilidad.
3. **Innovación en Criterios de Recomendación para la Empleabilidad:** Implementar criterios de recomendación innovadores que se integren en el sistema de recomendación, utilizando técnicas avanzadas de análisis de datos para personalizar las sugerencias de desarrollo profesional. Estos criterios estarán orientados a optimizar la relevancia y aplicabilidad de la formación académica, incrementando así las oportunidades de empleo de los graduados.

7. Justificación y Aplicación de la Metodología

Entendiendo los nuevos retos de la educación y su necesidad para que sus estudiantes acorten las brechas del cambiante mercado laboral que enfrenta frente a la nueva ola de industrialización 4.0, las áreas encargadas de la empleabilidad y desarrollo profesional deben reinventarse y establecer una hoja de ruta para poder enfrentar el voluble mercado laboral ecuatoriano Actual.

Para realizar una selección eficiente y acertada de las bases y metodologías a utilizar se plantea usar una metodología de Balance Score Card (BSC), esto se debe a su versatilidad y eficacia como herramienta de gestión estratégica. Quintero-Beltrán y Osorio-Morales (2018) analizan la relevancia del BSC en empresas en crisis, resaltando su utilidad para realinear estrategias y mejorar la toma de decisiones (Quintero-Beltrán, Osorio-Morales, 2018). Vega, Falcón y Jácome (2020) discuten el BSC como una herramienta esencial para la gestión del rendimiento organizacional, señalando su capacidad para actualizar y limitar estrategias en un entorno cambiante (Vega, Falcón & Jácome, 2020). Como se observa el uso de un BSC se necesita tener claro cuáles son los obstáculos y retos que enfrenta la organización, en este caso las instituciones de educación superior se enfrentan a una situación donde las mallas tradicionales no son suficiente para la inserción laboral de los estudiantes con la nueva tendencia de búsqueda de habilidades blandas y nuevas tecnologías. Para lo cual se plantean el siguiente BSC para una universidad privada ecuatoriana¹.

7.1. Selección de la base de datos

Tras la revisión del Balance Score Card del área de la institución privada de educación superior se establece que se existen dos fuentes de información primordiales para las áreas de relaciones corporativas y la de desarrollo profesional donde se debe utilizar bases propias de la organización y bases públicas.

¹ Véase el Anexo 1

7.1.1. Bases de datos para la creación de relaciones corporativas

Las bases que se han seleccionado para el análisis de las relaciones estratégicas de la universidad, del lado de información interna se ha seleccionado un reporte de la plataforma de *Carrer Services* de la universidad de donde se obtiene un reporte de las interacciones y solicitudes de puestos de trabajo de las empresas hacia la institución. Esta base se limita a un periodo anual referente al 2022-2023, donde también se encuentran los datos de identificación de las empresas y su periodicidad de oferta, la base de datos cuenta con la característica de contar con 10 columnas y un total de 7222 registros. Mientras que las bases de datos publicas seleccionadas fueron obtenidas de la Superintendencia de Compañías del Ecuador, siendo la primera el ranking empresarial creado por la entidad de control ecuatoriano, donde se toman en cuenta indicadores financieros y de talento humano para su creación,obteniendo un total de 22 columnas y 103339 entradas (Superintendencia de Compañías, 2023), en segunda instancia se ha seleccionado el directorio de empresas registradas en la Superintendencia de Compañías, donde uno de los factores mas relevantes es la clasificación de las empresas basadas en la Clasificación industrial internacional uniforme (CIIU) de todas las actividades económicas, esta fuente de información entrega un total de 198023 registros y 23 columnas entradas (Superintendencia de Compañías, 2023).

7.1.2. Bases de datos para la aplicación de microaprendizaje en el desarrollo profesional de estudiante

Para el objetivo de crear mejores herramientas para el micro aprendizaje de los estudiantes frente a las ofertas laborales publicadas en el portal de *Carrer Services* se obtuvo un reporte del portal en el cual se encuentra el tipo de empleo y la descripción del puesto obteniendo un total de 5823 entradas y 14 columnas. La base publica de datos que se decidió utilizar por factores internos como también factores externos como el posicionamiento de la plataforma de micro learning en el mercado fue UDEMY, la cual entrega información en un formato *Coma Separated Values*

(.csv) con un total de 7 columnas y 44571 registros, donde además de presentar el nombre del curso presenta una breve descripción el rating y calificación (Kaggle,s.f).

7.2 Limpieza, preprocesamiento y transformación de datos

Una vez que las bases de datos fueron seleccionadas, se procede a escoger la herramienta mas apta para trabajar con estas bases. Como primera observación se encuentra que las bases de datos son heterogéneas en su formato, ya que se cuenta con 4 bases de datos en formato .xlsx (Excel) y una base de datos en formato csv, razón por la cual se debe escoger un programa y lenguaje de programación el cual nos permita leer ambos tipos de bases de datos, otro punto a tomar en cuenta es cuales son los objetivos del proyecto, en este caso es crear un sistema de recomendación de microaprendizaje para los estudiantes y un modelo que permita visibilizar las relaciones estratégicas a trabajar desde la universidad por lo cual nos encontramos que los modelos posibles a utilizar recaen en la rama de análisis prescriptivo y predictivo (Lepenioti et. al, 2020). Razones que apuntan a las herramientas de estadísticos que usen un lenguaje tipo R o Python para su tratamiento, por lo cual se ha selecciona a *Python*, en su entorno de ejecución de *Google Colaboratory* debido a su amigable interfaz y a sus extensiones que permiten conectar con diferentes plataformas de Google y externas para obtener la información y poder realizar publicaciones y actualizaciones de manera sencilla.

Tras escoger el lenguaje de programación y entorno de ejecución adecuado se procede a generar un análisis exploratorio de las bases donde el uso de librerías es primordial.

Las librerías seleccionadas fueron, *Pandas* una herramienta esencial en el análisis de datos, proporcionando estructuras de datos flexibles y métodos para manipular tanto tablas numéricas como series temporales. Su elemento más destacado, es el *DataFrame*, el cual permite el manejo eficiente de tablas de datos con columnas de

tipos variados, facilitando tareas de importación, limpieza, transformación y análisis exploratorio de datos (McKinney, 2010).

Otra de las librerías importantes es *NumPy*, la cual constituye el núcleo de la computación científica en *Python*, ofreciendo un objeto de matriz multidimensional y rutinas para operaciones rápidas de matrices, esenciales para cálculos numéricos eficientes y formando la base de otras bibliotecas científicas (Oliphant, 2006).

Mientras que *Matplotlib*, se destaca por su generación de gráficos en 2D, facilitando la producción de figuras de calidad de publicación en diversos formatos y entornos (Hunter, 2007). Finalmente, se usa *Seaborn*, esta librería es un bien complementario a *Matplotlib*, que permite al usuario ofrecer una interfaz de alto nivel para la creación de visualizaciones estadísticas atractivas y útiles, esencial para la exploración y comprensión de datos complejos (Waskom et al., 2020).

Teniendo en cuenta las herramientas a usar debemos entender la naturaleza de los datos para lo cual, el análisis estadístico de los presenta una clasificación de los datos en dos tipos principales: cuantitativos y cualitativos. Los datos cuantitativos contienen números y medidas. Se dividen en dos subtipos: los continuos, como la utilidad de una empresa, que puede tomar cualquier valor en un rango, y los discretos, como el número de empleados en una empresa, que son valores enteros (Zamani et al., 2024). Por otro lado, existen datos cualitativos describen características y se dividen en nominales, como la clasificación por CIU de las empresas, que son categorías sin orden, y ordinales, como la clasificación por tamaño de la empresa, que tienen un orden específico (Wong et al., 2024).

Entender estos tipos de datos es de suma importancia. Los datos cuantitativos permiten usar medidas estadísticas (media, mediana, máximos, mínimos, etc.), mientras que los cualitativos suelen utilizar herramientas como frecuencias o tablas (Leijon et al., 2024). Esta diferencia es clave para elegir cómo analizar y visualizar

los datos. Por ejemplo, usamos histogramas para cuantitativos y gráficos de barras para cualitativos (Mutula & Engairo, 2024; Avădănei et al., 2024).

Para el manejo de bases de datos, un aspecto que es de vital relevancia es asegurar la integridad y completitud de los datos. En varios foros se establece que para que una base de datos sea considerada completa y apta para un análisis riguroso, el porcentaje de valores perdidos no debe superar el 5%, Criterio que será utilizado para la limpieza y selección de las columnas a utilizar. Este umbral es significativo porque garantiza la calidad y fiabilidad de los datos, elementos fundamentales para obtener resultados precisos y confiables en cualquier análisis estadístico. Además, mantener los valores perdidos por debajo de este límite facilita la aplicación de técnicas de imputación que permitan mantener la estructura y significado original de los datos evitando sesgos significativos (Brusil,2024).

Finalmente se presentan las tablas con la descripción de los datos de cada una de las bases de datos.

7.2.1 Preprocesamiento y transformación de las bases de datos para la creación de relaciones corporativas

Tabla 1: Base de datos del Carrer Services del directorio de empresas registrado en la plataforma

| # | Variable | Naturaleza Principal | Subdivisión |
|----------|--------------------------|-----------------------------|--------------------|
| 0 | unnamed:_0 | Cuantitativa | Discreta |
| 1 | organization_name | Cualitativa | Nominal |
| 2 | primary_contact | Cualitativa | Nominal |
| 3 | contact:_employer(count) | Cuantitativa | Discreta |
| 4 | email | Cualitativa | Nominal |
| 5 | phone | Cualitativa | Nominal |
| 6 | created | Cualitativa | Nominal |

Elaborado por: Carlos Brusil

Fuente: Plataforma de Carrer Service

Tabla 2: Base de datos del Ranking Empresarial 2023

| # | Variable | Naturaleza Principal | Subdivisión |
|----|----------------------------------|----------------------|-------------|
| 0 | posición | Cuantitativa | Discreta |
| 1 | año | Cuantitativa | Discreta |
| 2 | posición_ | Cuantitativa | Discreta |
| 3 | año.1 | Cuantitativa | Discreta |
| 4 | expediente | Cuantitativa | Discreta |
| 5 | nombre | Cualitativa | Nominal |
| 6 | tipo_compañía | Cualitativa | Nominal |
| 7 | actividad_económica | Cualitativa | Nominal |
| 8 | región_ | Cualitativa | Nominal |
| 9 | provincia | Cualitativa | Nominal |
| 10 | ciudad | Cualitativa | Nominal |
| 11 | tamaño | Cualitativa | Nominal |
| 12 | sector | Cualitativa | Nominal |
| 13 | cant._empleados | Cuantitativa | Discreta |
| 14 | activo_2022 | Cuantitativa | Continua |
| 15 | patrimonio_2022 | Cuantitativa | Continua |
| 16 | ingreso_por_ventas_2022 | Cuantitativa | Continua |
| 17 | utilidad_antes_del_impuesto_2022 | Cuantitativa | Continua |
| 18 | utilidad_del_ejercicio_2022 | Cuantitativa | Continua |
| 19 | utilidad_neta_2022 | Cuantitativa | Continua |
| 20 | ir_causado_2022 | Cuantitativa | Continua |
| 21 | ingreso_total_2022 | Cuantitativa | Continua |

Elaborado por: Carlos Brusil

Fuente: Superintendencia de Compañías, 2023

Las bases que se presentan en las tablas de la parte superior la interrelación para el posterior análisis se realiza una homologación de los nombres de las empresas a través de las columnas que llevan la información del nombre de las empresas,

estas tienen la particularidad que dentro de la plataforma de Carrer Services que esta columna se ingresa de forma manual por lo cual no son directamente homologable por errores de digitación o entradas erróneas por los diferentes participantes de la fase de ingreso de la información. Esto presenta un gran reto para el cual se utilizó un proceso de tratamiento para homologar las bases mediante el nombre de la organización(Brusil,2024).

Tras el proceso de homologación se genera la base obtenida mediante la unión de las diferentes bases con lo cual se obtiene la siguiente base que tiene un total de 54 columnas y 1907 entradas².

Esta base contiene valores repetidos por lo cual se procede a limpiar la base de estos valores con la finalidad de poder trabajar con una base simplificada. De manera que contamos con 28 columnas 1907 entradas (Brusil,2024).

Tabla 3: Base final después de homologación

| # | Variable | Naturaleza Principal | Subdivisión |
|----|--------------------------|----------------------|-------------|
| 0 | contact:_employer(count) | Cuantitativa | Discreta |
| 1 | ruc | Cuantitativa | Discreta |
| 2 | fecha_constitucion | Cualitativa | Nominal |
| 3 | tipo | Cualitativa | Nominal |
| 4 | país | Cualitativa | Nominal |
| 5 | región | Cualitativa | Nominal |
| 6 | representante | Cualitativa | Nominal |
| 7 | cargo | Cualitativa | Nominal |
| 8 | capital_suscrito | Cualitativa | Nominal |
| 9 | ciiu_nivel_1 | Cualitativa | Nominal |
| 10 | ciiu_nivel_6 | Cualitativa | Nominal |
| 11 | posición | Cuantitativa | Discreta |

² Véase anexo 2.

| | | | |
|----|----------------------------------|--------------|----------|
| 12 | año | Cuantitativa | Discreta |
| 13 | pocisión_ | Cuantitativa | Discreta |
| 14 | año.1 | Cuantitativa | Discreta |
| 15 | tipo_compañía | Cualitativa | Nominal |
| 16 | actividad_económica | Cualitativa | Nominal |
| 17 | tamaño | Cualitativa | Nominal |
| 18 | sector | Cualitativa | Nominal |
| 19 | cant._empleados | Cuantitativa | Discreta |
| 20 | activo_2022 | Cuantitativa | Continua |
| 21 | patrimonio_2022 | Cuantitativa | Continua |
| 22 | ingreso_por_ventas_2022 | Cuantitativa | Continua |
| 23 | utilidad_antes_del_impuesto_2022 | Cuantitativa | Continua |
| 24 | utilidad_del_ejercicio_2022 | Cuantitativa | Continua |
| 25 | utilidad_neta_2022 | Cuantitativa | Continua |
| 26 | ir_causado_2022 | Cuantitativa | Continua |
| 27 | ingreso_total_2022 | Cuantitativa | Continua |

Elaborado por: Carlos Brusil

Fuente: Superintendencia de Compañías,2023

7.2.2 Preprocesamiento y transformación de datos para la aplicación de microaprendizaje en el desarrollo profesional de estudiante

Tabla 4: Base de datos de plataforma de los registros de las postulaciones anuales 2022-2023

| # | Variable | Naturaleza Principal | Subdivisión |
|---|----------------------------|----------------------|-------------|
| 0 | Identificación | Cualitativa | Nominal |
| 1 | Cargo | Cualitativa | Nominal |
| 2 | Tipo de puesto | Cualitativa | Nominal |
| 3 | Empleador | Cualitativa | Nominal |
| 4 | Especializaciones deseadas | Cualitativa | Nominal |

| | | | |
|----|-------------------------------------|--------------|----------|
| 5 | Fecha de publicación | Cuantitativa | Discreta |
| 6 | Fecha de finalización | Cuantitativa | Discreta |
| 7 | Solicitud no OCR: Puesto(contar) | Cuantitativa | Discreta |
| 8 | Vistas únicas de los estudiantes | Cuantitativa | Discreta |
| 9 | Total, de vistas de los estudiantes | Cuantitativa | Discreta |
| 10 | Conteo de vistas de estudiantes | Cualitativa | Nominal |
| 11 | Aprobado | Cualitativa | Nominal |
| 13 | Descripción | Cualitativa | Nominal |

Elaborado por: Carlos Brusil

Fuente: Plataforma de Carrer Service

Tabla 5: Base datos de cursos UDEMY 2023

| # | Variable | Naturaleza Principal | Subdivisión |
|---|------------|----------------------|-------------|
| 0 | Titulo | Cualitativa | Nominal |
| 1 | Resumen | Cualitativa | Nominal |
| 2 | Enrollment | Cuantitativa | Discreta |
| 3 | Estrellas | Cuantitativa | Continua |
| 4 | Rating | Cuantitativa | Discreta |
| 5 | Link | Cualitativa | Nominal |
| 6 | Categoría | Cualitativa | Nominal |

Elaborado por: Carlos Brusil

Fuente: Kaggle, s.f.

La Tabla 1 muestra la información que obtiene la plataforma de las postulaciones realizadas para el proceso de reclutamiento. Esta tabla contiene información esencial como lo es la descripción del puesto, sin embargo, la descripción al estar en español se debe traducir al idioma ingles para que se pueda crear una interrelación con la base de los cursos UDEMY (véase Tabla 2). Para este proceso se utiliza el entorno de *Google Sheets* la cual tiene la función de traducción mediante la formula = *GOOGLETRANSLATE(Celda a traducir;"idioma de Origen";" idioma de Salida")*, con lo cual la tabla final se muestra de la siguiente manera.

Tabla 6: Base del Carrer Service transformada

| # | Variable | Naturaleza Principal | Subdivisión |
|----|-------------------------------------|----------------------|-------------|
| 0 | Identificación | Cualitativa | Nominal |
| 1 | Cargo | Cualitativa | Nominal |
| 2 | Tipo de puesto | Cualitativa | Nominal |
| 3 | Empleador | Cualitativa | Nominal |
| 4 | Especializaciones deseadas | Cualitativa | Nominal |
| 5 | Fecha de publicación | Cuantitativa | Discreta |
| 6 | Fecha de finalización | Cuantitativa | Discreta |
| 7 | Solicitud no OCR: Puesto(contar) | Cuantitativa | Discreta |
| 8 | Vistas únicas de los estudiantes | Cuantitativa | Discreta |
| 9 | Total, de vistas de los estudiantes | Cuantitativa | Discreta |
| 10 | Conteo de vistas de estudiantes | Cualitativa | Nominal |
| 11 | Aprobado | Cualitativa | Nominal |
| 12 | Descripción en Ingles | Cualitativa | Nominal |
| 13 | Descripción | Cualitativa | Nominal |

Elaborado por: Carlos Brusil

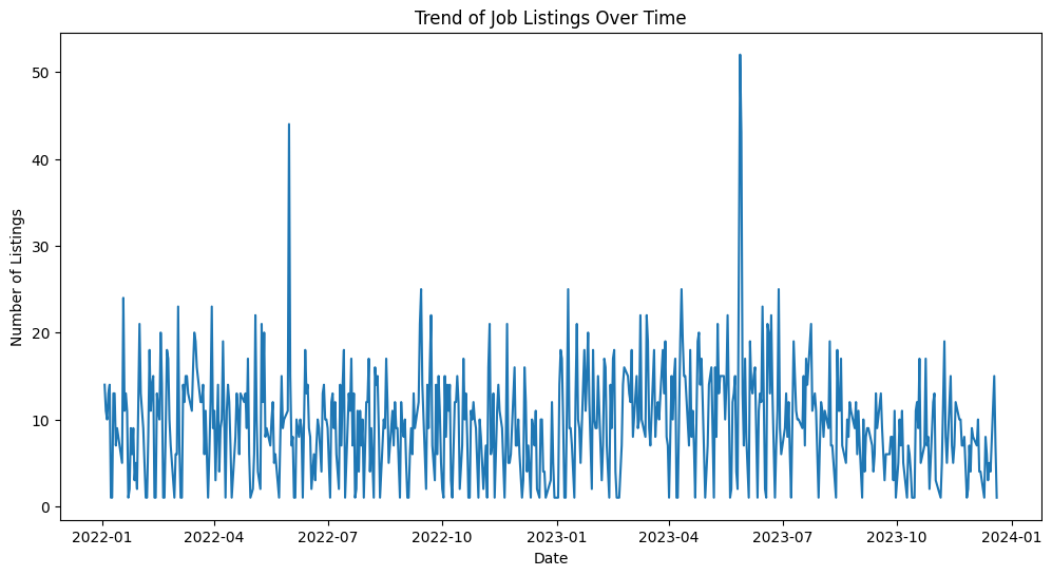
Fuente: Plataforma de Carrer Service

7.3. Identificación y descripción de variables

Entendiendo cuales han sido las bases finales y cuales son los datos que se posee procedemos a crear un análisis de la variable para generar un panorama general de la realidad que enfrenta la universidad, los *stakeholders* y proveedores de los diferentes servicios que se enfrentan en este proyecto (Brusil,2024).

Para comenzar debemos medir la afluencia de solicitudes que existe dentro de la plataforma por parte de los empleadores para lo cual se genera un grafico de serie de tiempo donde en el eje horizontal (eje x) encontraremos una línea temporal y el eje vertical (eje y) encontramos el numero de nuevos empleos ofertados (Brusil,2024).

Figura 1: Número de ofertas laborales durante el periodo 2022-2023



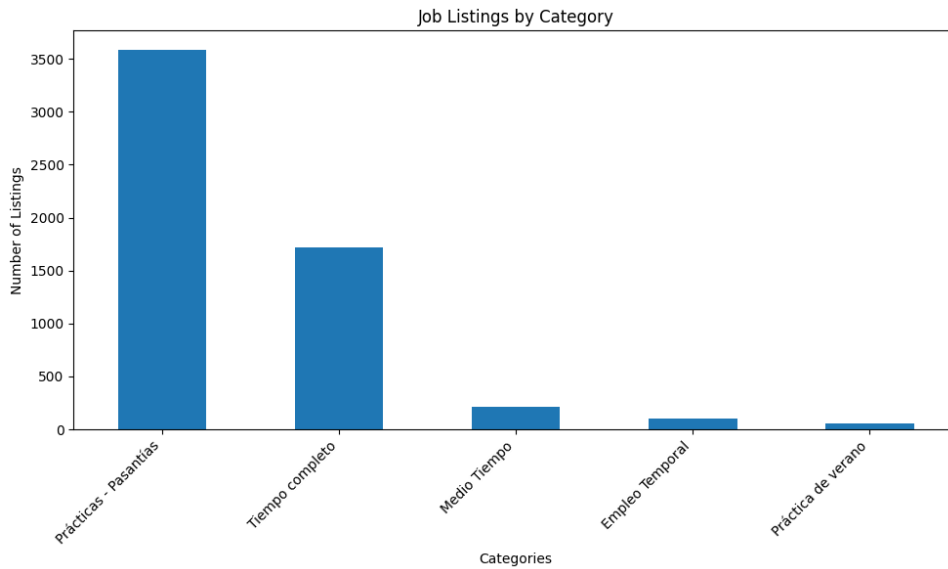
Elaborado por: Carlos Brusil

Fuente: Plataforma de Carrer Service

En este grafico se observa cierta estacionalidad de los datos ya que observamos que durante los periodos de los meses de abril y julio existe un alza de la demanda de puestos, esto se debe a varios factores como lo son las actividades dedicadas a la empleabilidad y el inicio del proceso de selección de pasantías universitarias por parte de las empresas para el semestre de septiembre.

Continuando con la revisión de la oferta laboral de la institución universitaria se procede a revisar cuales son los tipos de empleo que se ofrecen mediante la plataforma de *Carrer Service*.

Figura 2: Histograma del número de ofertas laborales por categoría.



Elaborado por: Carlos Brusil

Fuente: Plataforma de Carrer Service

En la Figura 2 podemos observar que una de las principales necesidades y uso de las empresas para la plataforma de *Carrer Services* es para encontrar pasantes lo cual es coherente con la filosofía y metas del área, también se resalta una interesante oportunidad ya que la siguiente categoría mas solicitada es la de empleos a tiempo completo lo cual puede sugerir que la herramienta también es útil para los *alumni*, lo cual abre una puerta para futuros estudios.

Para finalizar el análisis de la oferta laboral solicitada se crea una nube de palabras para entender tanto las necesidades de las empresas como cual es su estrategia de posicionamiento como marca empleadora.

Figura 3: Nube de palabras basado en la descripción del empleo.



Elaborado por: Carlos Brusil

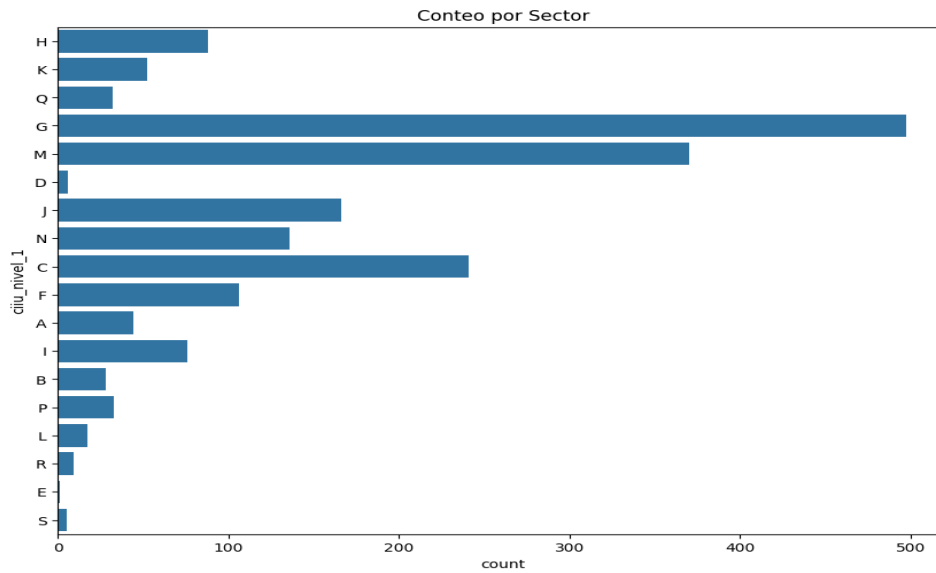
Fuente: Plataforma de Career Service

De la nube de palabras vemos que existe una alta importancia para las empresas de desarrollar las ideas de su empresa y la importancia de las carreras solicitadas, sin embargo, el gráfico deja bastantes dudas frente a cuáles son los intereses en sus habilidades, estas se muestran heterogéneas y difusas debido a la diversidad que pueden necesitar las organizaciones. Sin embargo, se muestran necesidades comunes como el trabajo de equipo, gestión de tiempo, el trabajo con clientes. También podemos identificar que de las carreras más solicitadas se encuentra el área de marketing, administración y sistemas. Finalmente vemos que las palabras con mayor frecuencia relacionadas al cargo son las pasantías, lo cual mantiene coherencia con la figura 2.

A continuación, se muestra la siguiente gráfica donde realiza un conteo de las empresas por su sector según el CIIU al primer nivel³.

³ Véase anexo 3

Figura 4: Distribución de las empresas por sector CIU nivel 1



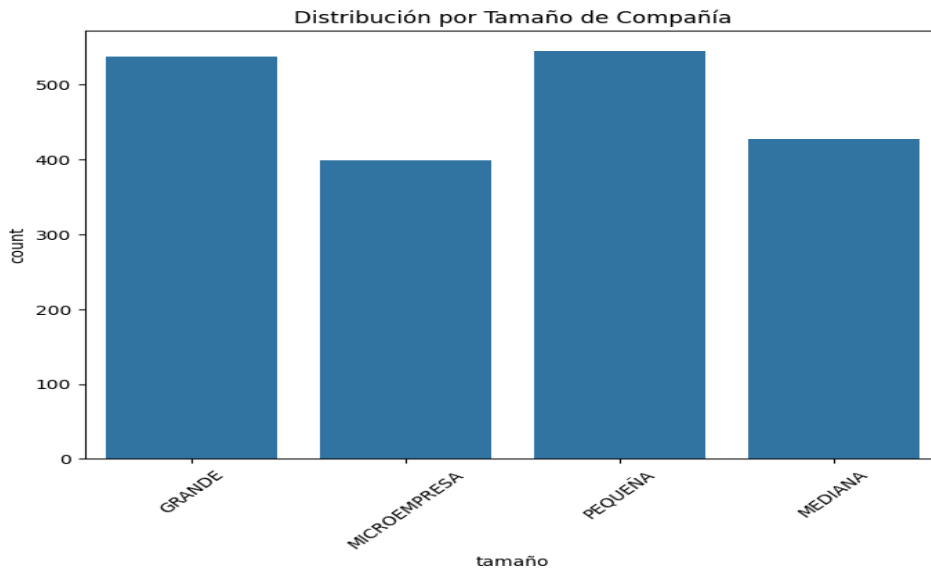
Elaborado por: Carlos Brusil

Fuente: Plataforma de Carrer Service y Superintendencia de Compañías

Dentro de este grafico se observa que existe tres industrias principales son aquellas que se dedican a comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas, actividades profesionales, científicas y técnicas, e industrias manufactureras.

Otro punto importante para resaltar en el análisis de las variables es el tamaño de empresas que se encuentran en la oferta laboral ya que estas deben estar alineadas a la estrategia del área, esta distribución se muestra en el siguiente gráfico.

Figura 5: Distribución por tamaño de la empresa



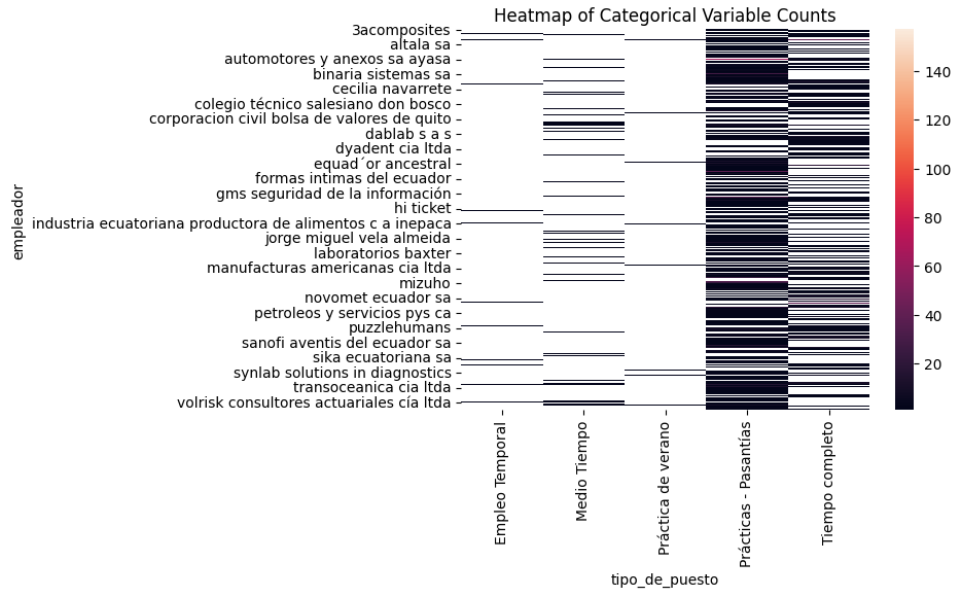
Elaborado por: Carlos Brusil

Fuente: Plataforma de Carrer Service y Superintendencia de Compañías

Se puede observar que la universidad tiene una distribución de tamaños de empresas bastante equilibrado siendo las empresas grandes y pequeñas las principales reclutadoras de nuevos talentos seguidos por las medias y finalmente, las microempresas.

Para entender mejor la demanda laboral continuamos con la densidad de solicitudes por empresas para lo cual se presenta la figura continuación en la cual podemos observar que cantidad de solicitudes encontramos por empresa y se resalta las más importantes.

Figura 6: Mapa de calor sobre la cantidad de solicitudes por empleador y tipo de empleo



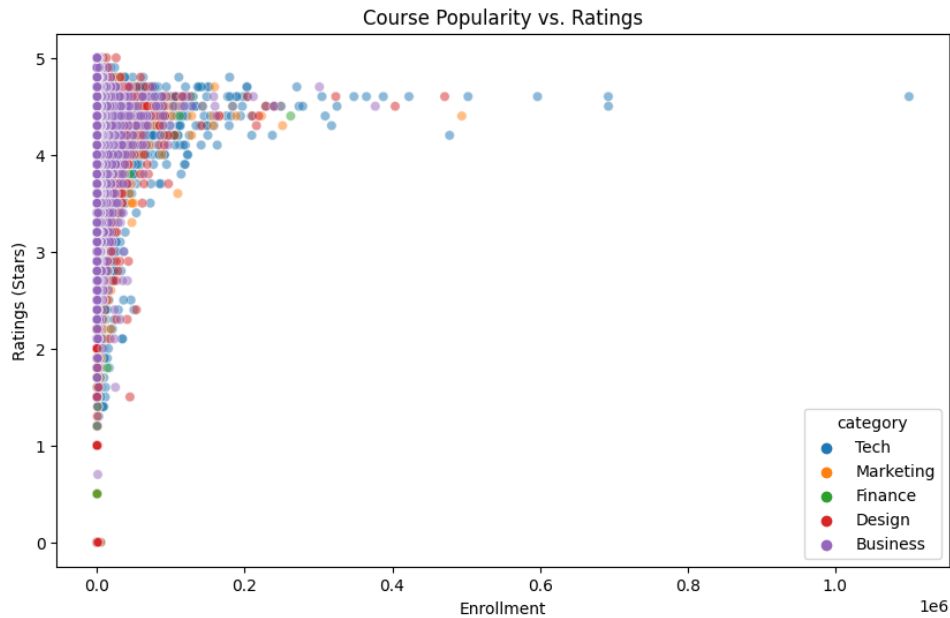
Elaborado por: Carlos Brusil

Fuente: Plataforma de Carrer Service

La figura demuestra la importancia de las pasantías, destacando algo muy importante y es que las empresas no publican mas de 20 postulaciones para una pasantía.

Para cerrar con este capítulo del proyecto se presenta cual es la distribución de los cursos de micro aprendizaje de UDEMY según sectores de conocimiento y su calificación.

Figura 7: Clasificación de los cursos por su área de conocimiento, enrollment y calificación



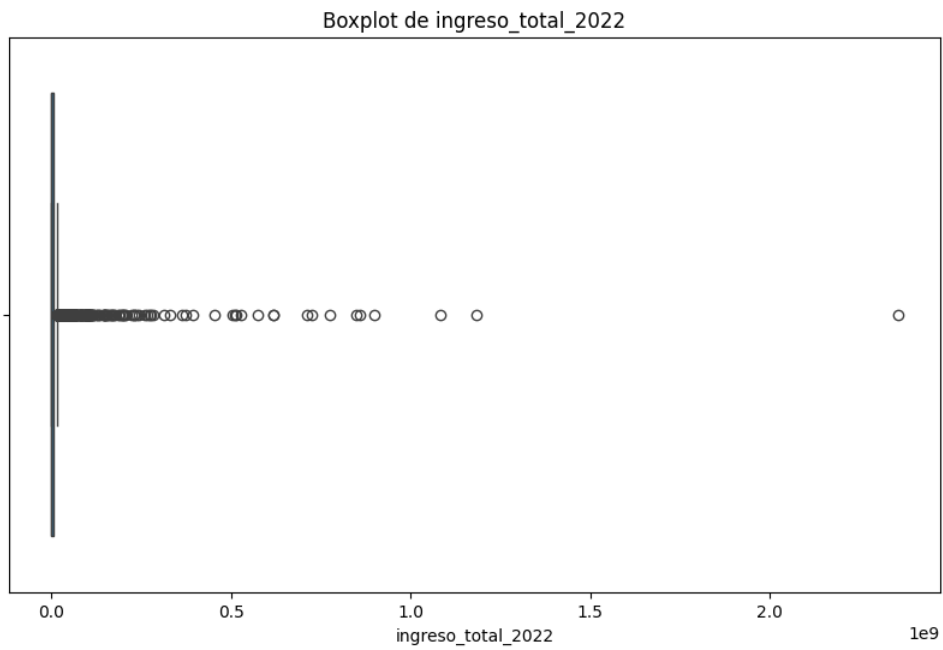
Elaborado por: Carlos Brusil

Fuente: Kaggle

En la oferta académica de micro aprendizaje para la preparación de los estudiantes podemos observar que la mayor oferta es en el área de conocimiento de negocios que coincide con la demanda laboral presente en las publicaciones de la plataforma. Lo que indica que la oferta académica de UDEMY es útil para la preparación de los estudiantes.

Finalmente realizamos una revisión de las variables cuantitativas. Para esta revisión se realiza un grafico de bigotes para observar la distribución y entender si es necesario realizar procesos en las variables.

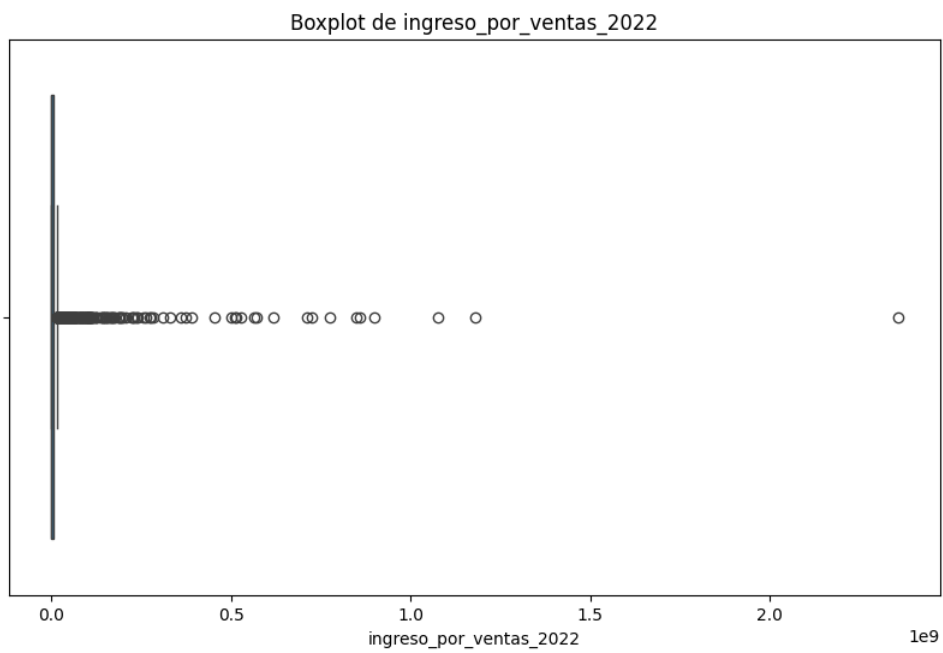
Figura 8: Grafico de bigotes



Elaborado por: Carlos Brusil

Fuente: Superintendencia de Compañías

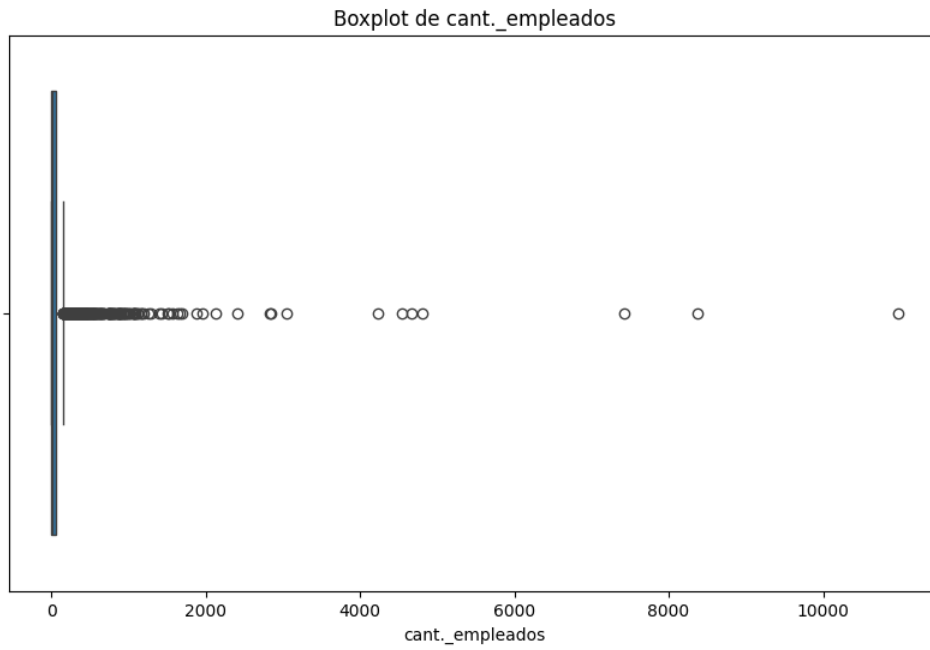
Figura 9: Grafico de bigotes



Elaborado por: Carlos Brusil

Fuente: Superintendencia de Compañías

Figura 10: Grafico de bigotes

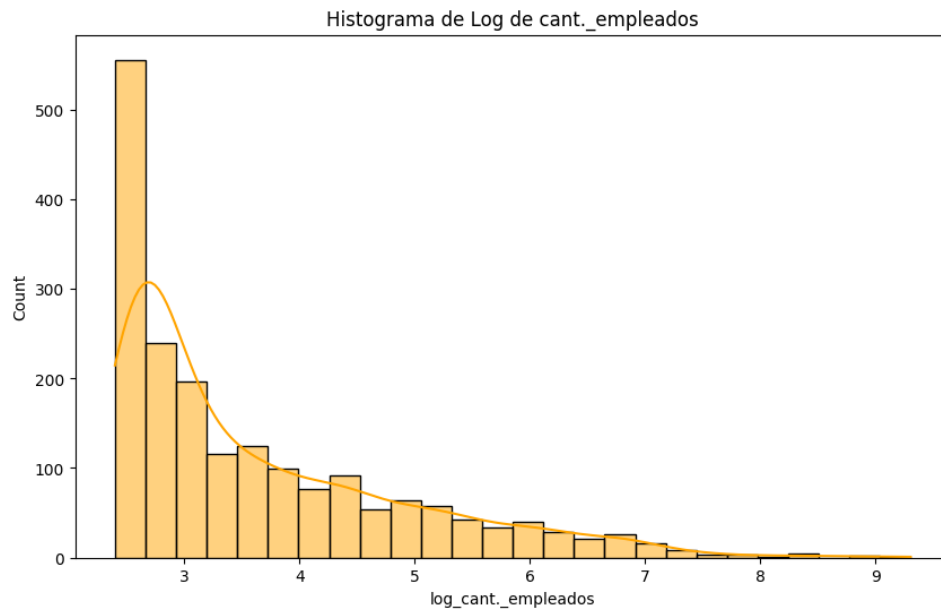


Elaborado por: Carlos Brusil

Fuente: Superintendencia de Compañías

Tras la revisión de estos gráficos observamos que no existe una normalidad en la distribución de los datos por lo cual se realiza una transformación logarítmica para corregir este problema, como se muestra en los siguientes gráficos.

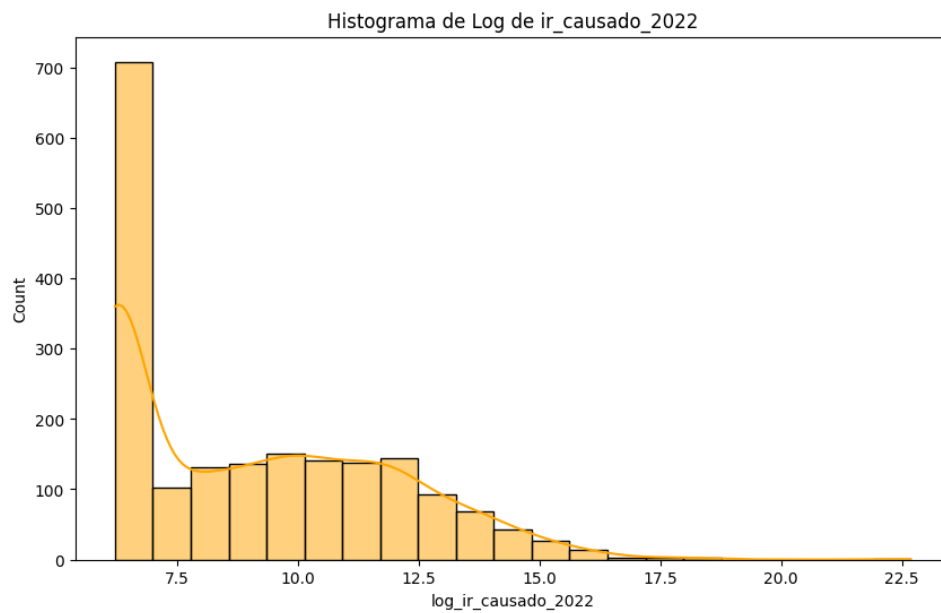
Figura 11: Transformación logarítmica de la variable empleados



Elaborado por: Carlos Brusil

Fuente: Superintendencia de Compañías

Figura 12: Transformación logarítmica del impuesto a la renta causado



Elaborado por: Carlos Brusil

Fuente: Superintendencia de Compañías

7.4. Selección de los modelos

El presente apartado tiene como objetivo entender la importancia de la aplicación de los distintos modelos matemáticos y estadísticos para el cumplimiento de los objetivos del estudio. Los modelos actuales son apoyados por las nuevas tecnologías de la información y la comunicación, por lo cual requieren de una labor eficiente en la organización y desarrollo de la investigación científica (Bowen, 1997).

El vínculo entre la teoría y la aplicación práctica de los modelos se ha fortalecido a través de su inclusión en programas de postgrado y pregrado en diversas universidades, subrayando la importancia de una selección adecuada de modelos estadísticos para la investigación científica y la toma de decisiones basada en datos (Guerra et al., 1993).

El proyecto tiene como objetivo mejorar la empleabilidad de los estudiantes universitarios a través del desarrollo de estrategias integradas en relaciones externas y programas de desarrollo profesional, la aplicación de modelo como, lo es el clustering emerge como un enfoque estratégico para optimizar la gestión de relaciones con empresas y organizaciones relevantes.

La metodología propuesta incluye la implementación y comparación de dos modelos de clustering distintos para mejorar la gestión de las relaciones estratégicas, la optimización de estrategias de colaboración mediante criterios de clustering. También se propone el uso de un sistema de recomendación con el fin de cerrar la brecha entre las habilidades y competencias adquiridas por los estudiantes y las demandas del mercado laboral contemporáneo, sino también asegurar que los graduados estén mejor preparados para enfrentar desafíos y aprovechar oportunidades en sus campos profesionales.

Este enfoque refleja un compromiso con la aplicación práctica y orientada al análisis de datos para mejorar la inserción profesional de los graduados, situando a la

modelización estadística en el centro de las estrategias de desarrollo profesional y de relaciones externas en el ámbito universitario.

7.4.1 Clustering

Para poder cumplir con el objetivo de gestionar relaciones corporativas, se seleccionan los modelos de clasificación de aprendizaje no supervisado conocidos como *data clusterings*, estos algoritmos describen el proceso de clasificar los datos en grupos que tienen un alto grado de similitud mientras que se distancia entre los otros grupos de datos (Niknam & Amirí, 2009).

Los modelos de clasificación se dividen en dos subtipos de modelización: el primero se lo conoce como clustering que tiene como objetivo maximizar la separación de las clases o subpoblaciones, esto se realiza encontrando la mejor partición de las instancias, mediante la instalación de hiperplanos que dividen el espacio m-dimensional. El siguiente modelo de la clasificación es el modelo mixto el cual no busca maximizar la separación de clases, este modelo de *clustering* se enfoca en el grado de pertenencia que tiene cada clase, por lo cual se le asigna a una o más clases, esta técnica de modelización es comúnmente probabilística, pero también puede ser de carácter Bayesiano (Davidson, 2002).

El proyecto ha seleccionado diferentes variables las cuales son de naturaleza heterogénea, de los cuales destacan las variables numéricas discretas y continuas, mientras que las variables cualitativas son de carácter discretas y nominales. Para lo cual se ha seleccionado las siguientes variables 'log_contact:_employer(count)', 'log_activio_2022', 'log_ir_causado_2022', 'log_cant._empleados', 'cambio_posicion', 'ciiu_nivel_1', 'tamaño'.

Para observar cual es el mejor modelo que se adapta se utilizara un algoritmo *K-Means*, el cual se inicia con la elección aleatoria de lo centroides (k), y se continua con la asignación de cada punto a uno de los centroides mas cercanos, basando la

decisión en la distancia euclidiana como una media de proximidad (Romanuke, 2023; Wibowo et al., 2023). La cual sigue la ecuación 7.1.

$$(7.1) d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Tras realizar la asignación de los datos a los centroides se realiza un recalcu de los centroides como un promedio de todos los puntos asignados al clúster, este paso pasa a formar parte de un proceso iterativo que termina cuando los centroides dejan de presentar cambios significativos. Lo que indica que el algoritmo ha convergido (Kumara et al., 2019; Cabrera et al., 2019, Brusil, 2024).

El segundo tipo de modelo a utilizar es una derivación del modelo antes presentado y conocido como *DBSCAN*, abreviatura que nace del inglés Density-Based Spatial Clustering of Applications with Noise, una técnica de clustering que destaca por su capacidad para identificar clústeres o clases de una forma arbitraria, lo que permite diferenciar a los modelos significativamente de métodos basados en centroides que presuponen formas esféricas de clústeres (Ester, et al., 1996). Este algoritmo analiza la densidad de los puntos de datos, agrupando los datos por su medida de dispersión, mientras identifica y excluye a los datos aislados como outliers. La eficacia de los resultados del algoritmo DBSCAN radican en su flexibilidad para tratar con estructuras de clústeres desconocidas o irregulares, lo que permite una detección más intuitiva y exacta de patrones en conjuntos de datos complejos (Ester et al., 1996; Brusil, 2024).

Debido al enfoque basado en la densidad del algoritmo DBSCAN, se requieren dos parámetros fundamentales: *eps* y *min Pts*. El parámetro *eps* determina la distancia máxima del *neighborhood*, mientras que *min Pts* es el número mínimo de puntos

requeridos dentro de un radio ϵ para formar un clúster densamente poblado (Ester et al., 1996). Estos parámetros permiten al DBSCAN adaptarse a diferentes densidades y tamaños de clusters, proporcionando una flexibilidad significativa en comparación con métodos de clustering más rígidos.

La clasificación dentro del modelo DBSCAN sigue reglas específicas, donde un punto p solo se considera central si tiene, al menos, $\min Pts$ puntos dentro de su radio ϵ , lo que indica una alta densidad local (Schubert, et al., 2017). Si un punto q es directamente alcanzable desde p , cuando está dentro del radio ϵ de q , y p es un punto central. El rango de alcance del algoritmo se extiende a través de una cadena de puntos directamente alcanzables, lo que permite la formación de clusters basados en la proximidad y densidad de puntos, diferenciando claramente los clusters de las anomalías o ruido existente en el modelo esto se lo puede observar en la fórmula 7.2.

*(7.2) donde p es un punto central
y q es un punto alcanzable
cuando hay una secuencia p_1, \dots, p_n
con $p_1 = p$ y $p_n = q$
donde cada p_{i+1} es alcanzable desde p_i*

Cuando empieza el proceso iterativo de construcción de clústeres en el método DBSCAN, este comienza con un punto no visitado, evaluando si cumple con los criterios para ser considerado un centroide y, de cumplir las condiciones, se continúa formando un clúster a su alrededor. Este clúster se expande encontrando todos los puntos alcanzables bajo las reglas establecidas en los párrafos anteriores, enfoque que facilita la identificación efectiva de clústeres en presencia de ruido y la clasificación de puntos no alcanzables como anomalías (Rahmah & Sitanggang, 2016). La iteración seguirá hasta que todos los puntos hayan sido examinados,

permitiendo revelar la estructura de clústeres subyacente sin la necesidad de especificar un número previo de clústeres.

7.4.2 Sistema de recomendación

Cuando se habla de cerrar la brecha entre las habilidades y competencias adquiridas por los estudiantes y las demandas del mercado laboral contemporáneo, se opta por los sistemas de recomendación basados en contenido, estos desempeñan un papel esencial en la filtración y personalización de la vasta cantidad de información accesible para los usuarios de las diferentes plataformas virtuales. Estos sistemas sugieren productos o servicios a los usuarios con base en la descripción de estos y en los intereses del usuario, aplicándose en una amplia gama de dominios como sitios web, artículos de noticias, restaurantes y productos a la venta. El eje central de estos sistemas es el desarrollo de un perfil de usuario que se actualice de manera persistente con base en la retroalimentación sobre los artículos sugeridos, lo cual facilita una personalización más detallada de las recomendaciones para el usuario (Pazzani & Billsus, s.f.).

Una de las diferencias más destacadas que se encuentra en los modelos de recomendación radica entre el filtrado colaborativo y el basado en contenido. El primero tiene su raíz en las interacciones entre usuarios y artículos para formular recomendaciones, mientras el segundo dirige su atención a las características de los artículos y en las preferencias explícitas o implícitas del usuario en este caso en las preferencias al escoger a su empleador. A pesar de los beneficios que presenta el método de filtrado basado en contenido, este enfrenta retos como la tendencia a la sobre especialización y la creciente necesidad de recolectar información cada vez más detallada sobre las preferencias de los usuarios (Thorat, Ashwini & Seema, 2023).

Además, los sistemas de recomendación basados en contenido tienen la oportunidad de beneficiarse de enfoques contextuales, considerando diversos

factores como la ubicación del usuario, el momento y la compañía para efectuar recomendaciones más adecuadas (Javed et al., 2021).

Existen evaluaciones experimentales de distintos enfoques de recomendación, hecho que ha permitido demostrar que la implementación de modelos basados en espacio vectorial, aprendizaje profundo y modelos semánticos en sistemas de recomendación basados en contenido tienen alta efectividad. En particular, existe una aplicación conocida como BERT para el modelado del lenguaje que ha probado ser eficaz en el ámbito de las recomendaciones de películas (Rodríguez-Hernández et al., 2020).

Una de las formas de sortear los obstáculos habituales de los sistemas de recomendación como la escasez de datos y la sobre especialización, que se demuestra que han incrementado la precisión y la robustez de las recomendaciones, son la selección de características específicas para el objetivo del proyecto y la adopción de métodos de filtrado multiatributo en sistemas de recomendación basados en contenido pueden solucionar problemas (Son & Kim, 2017).

Para tomar la decisión entre utilizar un sistema de recomendación filtrado basado en contenido o colaborativo se debe a varios factores como lo pueden ser el tipo de producto o servicio y las preferencias del consumidor. Para productos de experiencia, como lo pueden ser películas, música o cursos de educación continua, los consumidores pueden preferir recomendaciones que se alineen más estrechamente con sus gustos personales, mientras que, para productos de búsqueda, el filtrado colaborativo puede resultar en evaluaciones más adecuadas (Liao & Sundar, 2021). En el presente estudio el enfoque del modelo se enfoca en generar un modelo que permita preparar a los estudiantes para su paso a la empleabilidad en base a la necesidad de herramientas y habilidades de la empresa, en este método se desarrolla basado en las variables de resumen y descripción del

cargo permitiendo encontrar el curso ideal para los requerimientos de la empresa (Brusil, 2024).

8. Resultados

Tras una exhaustiva ejecución y análisis de los modelos seleccionados para cumplir con los objetivos del proyecto, se ha logrado obtener una serie de resultados significativos que nos permiten acercarnos al cumplimiento de las metas establecidas en los primeros capítulos.

Inicialmente el estudio se enfocará en la presentación de los resultados obtenidos a partir de los modelos de clasificación. Para evaluar el desempeño y la eficacia de estos modelos, se han empleado dos métodos distintos, los cuales han sido cuidadosamente seleccionados para asegurar una evaluación integral y precisa de los modelos en cuestión. A continuación, se presenta las pruebas seleccionadas y los resultados para su interpretación.

El primer método que se utiliza para medir la eficiencia de la clasificación de los modelos K-Means y DBSCAN es el uso de la métrica Silhouette Score, la cual mide qué tan bien se ha realizado el agrupamiento de los datos. Según Rousseeuw (1987), el Silhouette Score varía de -1 a 1, donde un valor más alto indica que los objetos están bien emparejados dentro de su propio clúster y mal emparejados con clústeres vecinos, facilitando así una interpretación clara de la calidad del clustering (Rousseeuw, 1987). Esta idea se ve reforzada por Han, Lamber, y Pei (2011), quienes destacan la importancia de utilizar métodos robustos y confiables en el análisis de datos complejos, ofreciendo un marco teórico sólido sobre los principios de minería de datos (Han, Lamber & Pei, 2011).

Para calcular matemáticamente el Silhouette Score para cada punto i , se realiza primero el cálculo de $a(i)$, que es la distancia promedio que existe de i a los demás puntos en el mismo clúster, parámetro que mide la cohesión.

$$(8.1) \ a(i) = \frac{1}{|C|-1} \sum_{j \in C, j \neq i} d(i, j)$$

Se continúa calculando $b(i)$, que es la menor distancia de i a los puntos en todos los otros clusters, lo que mide la separación.

$$(8.2) \ b(i) = \min_{k \neq c} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j)$$

El Silhouette Score para el punto i se obtiene entonces como la diferencia entre $b(i)$ y $a(i)$ normalizada por el máximo de $a(i)$ y $b(i)$.

$$(8.3) \ s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Este enfoque matemático muestra el funcionamiento y utilidad del Silhouette Score en proporcionar una medida cuantitativa de la calidad del clustering. Cuando el valor del Silhouette Score se acerca al +1 sugiere que los clústeres están bien definidos y separados, indicando una asignación precisa de las muestras a sus respectivos clústeres (Kaufman & Rousseeuw, 1990). Esta característica es fundamental para asegurar que el modelo de clustering tenga una alta confiabilidad en la asignación de datos a clústeres específicos. Tan, Steinbach, y Kumar (2005) apoyan esta visión, ilustrando cómo la aplicación práctica del Silhouette Score en análisis de datos reales puede optimizar la agrupación de datos.

Por otro lado, un Silhouette Score cercano a 0 puede señalar que las muestras están en el límite entre dos clústeres, sugiriendo una posible superposición entre clústeres y, por ende, una ambigüedad en la asignación de clústeres (Kaufman & Rousseeuw, 1990). Este escenario plantea desafíos en la interpretación de los resultados del clustering, ya que indica una menor claridad en la separación entre grupos de datos. En este contexto, Xu y Wunsch (2005) discuten cómo la selección cuidadosa de

algoritmos de clustering y la evaluación de sus resultados mediante métricas como el Silhouette Score son esenciales para superar estos desafíos (Xu & Wunsch, 2005).

Un valor cercano a -1, en contraste, implica una posible asignación incorrecta de las muestras a los clústeres, lo que puede reflejar una elección inapropiada del número de clústeres o una ineficacia del algoritmo de clustering utilizado para ese conjunto de datos específico (Rousseeuw, 1987). Tales hallazgos son indicativos de la necesidad de reevaluar los parámetros del modelo o explorar diferentes algoritmos de clustering, una idea que Xu y Wunsch (2005) exploran al ofrecer una revisión exhaustiva de los algoritmos de clustering y sus aplicaciones (Xu & Wunsch, 2005).

Una vez que entendido como funciona la medida y la manera de interpretarlo se presenta los resultados de los modelos realizados para la gestión estratégica de relaciones corporativas.

Tabla 7: Resultado de los resultados Silhouette Score

| | |
|---------------------------|-------|
| Silhouette Score K-Means: | 0.31 |
| Silhouette Score DBSCAN: | -0.29 |

Elaborado por: Carlos Brusil

Fuente: Brusil, 2024

Estos resultados nos permiten inferir que el mejor modelo es el modelo K-Means que se acerca más al uno positivo permitiéndonos entender que el modelo es más eficiente tanto en la cohesión de cada uno de los puntos como en la separación. También permite decidir que el número eficiente de clústeres para los modelos son tres.

El segundo indicador que se toma en cuenta para medir la eficiencia de los modelos y seleccionar el más adecuado, se presenta al índice de Calinski-Harabasz, una

herramienta muy usada en el análisis de clústeres, lo cual la ha convertido en el foco de varios estudios recientes que buscan evaluar su eficacia y aplicabilidad en diferentes contextos. Por ejemplo, Lukasik et al. (2016) estudiaron la naturaleza de la aplicación del Algoritmo de Polinización de Flores en clustering, durante su investigación utilizaron el índice de Calinski-Harabasz como criterio de optimización, obteniendo resultados prometedores en comparación con el procedimiento K-Means (Lukasik, et al., 2016). Debido a este estudio se puede determinar la versatilidad del índice para evaluar algoritmos de clustering más allá de los métodos tradicionales.

Otras investigaciones como, la investigación de Ashari et al. (2023) sobre la clasificación de áreas afectadas por inundaciones en Yakarta utilizó el índice de Calinski-Harabasz para evaluar la agrupación, encontrando que la agrupación con 3 clústeres era el valor óptimo para clasificar (Ashari, et al., 2023).

El índice de Calinski-Harabasz se calcula matemáticamente se fundamenta en la evaluación de la dispersión, tanto intergrupala como externa, de los clústeres formados. Matemáticamente, se calcula a partir de la relación entre la dispersión entre los clústeres y la dispersión dentro de los clústeres, multiplicada por un factor que toma en cuenta el número total de muestras y el número de clústeres.

$$(8.4) b(i) = \frac{T_r(B_k)}{T_r(W_k)} \times \frac{N-k}{k-1}$$

De la ecuación 8.4 la matriz de dispersión trazada se simbolizada como, $T_r(B_k)$. La matriz representa la suma de cuadrados entre los grupos, y representa la variabilidad de los clústeres. Cuando el valor es alto en esta medida indica que los clústeres están bien separados entre sí.

Por otro lado, la matriz de dispersión dentro de los clústeres, denotada como $T_r(B_k)$, es la suma de cuadrados dentro de cada grupo. Esta medida evalúa qué tan

cercanas están las muestras dentro de un mismo clúster, y nos indica la densidad existente en el clúster.

El número total de muestras es representado por N , y el número de clústeres, se define por la letra k , son esenciales en el cálculo. El índice utiliza estos valores como ajuste de la dispersión, tomando en cuenta el tamaño del conjunto de datos y la cantidad de clústeres formados.

Para la interpretación del índice de Calinski-Harabasz debemos tener en cuenta que, si el valor es elevado, indica una estructura de clasificación óptima, caracterizada por una alta separación entre clústeres y una alta densidad dentro de los mismos (Lukasik, et al., 2016). Indicando que las muestras dentro de cada clúster sí comparten una alta similitud, mientras que hay claras distinciones entre diferentes clústeres.

El índice de Calinski-Harabasz es particularmente útil para comparar la efectividad de diferentes configuraciones de clustering. Al variar el número de grupos y calcular el índice para cada configuración, se puede determinar el número óptimo de clústeres que maximiza el índice de Calinski-Harabasz (Ashari et al., 2023). Una de las principales diferencias que existe entre el Silhouette Score y el índice Calinski-Harabasz, que se mueve dentro de un rango de -1 a 1, el índice no se mide en un rango definido, lo que permite una mayor flexibilidad en su interpretación para comparar diferentes métodos o configuraciones de clustering (Maulik & Bandyopadhyay, 2002).

Con la explicación matemática y de como se debe interpretar el índice, se procede a presentar los resultados de ambos modelos.

Tabla 8: Resultado de los resultados Calinski-Harabasz

| | |
|----------------------------|---------|
| Calinski-Harabasz K-Means: | 1714.67 |
|----------------------------|---------|

| | |
|---------------------------|-------|
| Calinski-Harabasz DBSCAN: | 24.61 |
|---------------------------|-------|

Elaborado por: Carlos Brusil

Fuente: Brusil, 2024

Dado la teoría y funcionamiento matemático del resultado se encuentra que el modelo K-Means para la clasificación de empresas es de mayor calidad que el obtenido del DBSCAN.

Finalizando el análisis de los resultados vamos a continuar midiendo los resultados del sistema de recomendación planteado para el fortalecimiento de los programas de desarrollo profesional.

En la evaluación de sistemas de recomendación, una métrica bien recibida es la precisión en las primeras K recomendaciones, conocida como Precisión K . Métrica ampliamente aceptada por la comunidad científica, permitiendo a los investigadores cuantificar la calidad de las recomendaciones proporcionadas a los usuarios, enfocándose específicamente en la relevancia de los ítems sugeridos dentro de las primeras K posiciones de la lista de recomendaciones (Herlocker, et al., 2004).

La Precisión K se define matemáticamente como la relación que existe entre el número de ítems relevantes encontrados en las primeras K posiciones dividido por el número total de ítems recomendados en esas K posiciones, proporcionando un indicador claro de la efectividad de un sistema de recomendación en términos de su capacidad para identificar ítems de interés para el usuario (Manning, Raghavan, & Schütz, 2008).

$$(8.5) \text{Precisión}(k) = \frac{\text{items relevantes en } k}{\text{total de items en las } k \text{ posiciones}}$$

El algoritmo de esta métrica implica iterar a través de un conjunto de recomendaciones, el cual verifica la relevancia de cada ítem recomendado con

respecto a las preferencias del usuario. Este proceso iterativo también refleja la eficiencia del modelo en clasificar estos ítems de manera que las recomendaciones más relevantes ocupen las primeras posiciones de la lista de sugerencias. La relevancia de un ítem, en este contexto, es definida por el investigador o desarrollador, mediante la evaluación entre el resumen o características del ítem y los intereses o necesidades a las que se pueda adaptar al usuario (Ricci, Rokach, & Shapira, 2011).

En conclusión, la Precisión K se establece como una herramienta altamente fiable en la evaluación de sistemas de recomendación, proporcionando un marco cuantitativo para medir la relevancia de las recomendaciones hechas a los usuarios. La importancia de esta métrica tiene su raíz en la capacidad para reflejar la calidad de las sugerencias más visibles para el usuario.

Para medir la precisión del presente modelo se realiza con la relevancia que toma el resumen del curso, donde se establece como factores a tomar en cuenta la valoración entregada por los usuarios anteriores y la concordancia de los títulos y el resumen con la recomendación. Tras la iteración que se realiza nuestro sistema de recomendación encuentra que las primeras 4 recomendaciones que se obtienen del sistema de recomendación tienen una precisión del 100%, pero a partir de la 5ta. Recomendación el sistema empieza a entregar recomendaciones menos precisas bajando su rendimiento hasta la décima donde la décima recomendación baja por debajo del 40%, que es el porcentaje mínimo de relevancia que se ha decidido se acepte en el estudio (Brusil, 2024).

9. Estrategia Organizacional

El punto donde se intercepta la educación superior y el mercado laboral contemporáneo es la empleabilidad de los graduados, la cual emerge como una prioridad crítica, lo que ha impulsado a las instituciones de educación a reevaluar y rediseñar sus programas y estrategias. El objetivo de la reevaluación y rediseño es cerrar la brecha entre las competencias adquiridas por los estudiantes durante su formación y las dinámicas necesidades del mercado laboral. El uso de tecnologías avanzadas como, el análisis de datos, y sus herramientas que para el presente trabajo son, los modelos de clustering y sistemas de recomendación, los cuales presentan una estrategia prometedora para abordar este desafío, permitiendo a las universidades, crear relaciones estratégicas alineadas a la actualidad, además de ofrecer una orientación profesional más personalizada y efectiva.

Cuando se habla de relaciones estratégicas, se ha enfocado la resolución de este problema mediante los modelos de clustering o clasificación, específicamente el método K-means comparado con DBSCAN, que han demostrado ser particularmente efectivos en la mejora de la gestión de relaciones con empresas y organizaciones externas. Ambos modelos fueron sometidos a diferentes indicadores como lo son en los índices de Silhouette score y Calinski-Harabasz, índices se ha identificado que el modelo K-means supera en rendimiento al DBSCAN en el contexto de formar alianzas estratégicas. Este hallazgo subraya la capacidad del modelo K-means para agrupar organizaciones de manera eficiente, facilitando la identificación de socios estratégicos que puedan ofrecer oportunidades de empleo alineadas con las competencias y aspiraciones de los graduados.

Además, de la optimización de estrategias de colaboración mediante criterios derivados de los modelos de clustering afinar sus alianzas con el sector real. El establecer criterios claros para la selección y mantenimiento de estas alianzas, es de vital importancia establecer medidas claras como lo son la cantidad de empleados de las organizaciones, su estabilidad para mantener los rankings, y su rendimiento financiero, mediante estos puntos las universidades pueden garantizar una correspondencia más estrecha entre las habilidades de los graduados y las

necesidades del mercado, mejorando así las oportunidades de prácticas profesionales y la empleabilidad general.

La innovación en nuestro modelo no se limita únicamente a la forma en que los estudiantes se relacionan con el contenido; se expande significativamente hacia la mejora de la orientación profesional proporcionada. A través de la implementación de sistemas de recomendación basados en contenidos, hemos logrado alcanzar una precisión del 75% después de las primeras 5 recomendaciones. Este resultado resalta el considerable potencial que las técnicas avanzadas de análisis de datos tienen para personalizar la orientación profesional de manera efectiva. No obstante, contextualizar el resultado es de suma importancia. Cuando se realizan las comparaciones, según Johnson y Lee (2022), la precisión promedio para sistemas similares en el sector es del 65%, lo que indica que nuestro modelo supera la norma establecida (Johnson & Lee, 2022).

Sin embargo, es importante abordar las posibles limitaciones del enfoque presentado. La mejora de la precisión del modelo en futuras iteraciones podría verse beneficiada por la consideración de sesgos presentes en los datos de entrenamiento y la diversificación de las recomendaciones (Lopez et al., 2024). Además, incorporar técnicas adicionales de evaluación, como la validación cruzada, podría proporcionar una perspectiva más robusta sobre la efectividad del modelo (Davis & Thompson, 2023).

Estos sistemas no solo incrementan la relevancia de las oportunidades de desarrollo profesional presentadas a los estudiantes, sino que también amplían las posibilidades de que los graduados encuentren empleo en campos que verdaderamente coinciden con sus intereses y habilidades. Este enfoque subraya la importancia de la personalización en la educación y orientación profesional, abriendo nuevas vías para mejorar la empleabilidad de los estudiantes universitarios.

La estrategia organizacional debe integrar estos avances tecnológicos no solo como una solución a los desafíos actuales, sino también como una inversión en el futuro de la educación superior y la empleabilidad de los graduados. La eficiencia de utilizar nuevas tecnologías, como los modelos de clustering y sistemas de recomendación, radica no solo en su capacidad para analizar grandes volúmenes de datos y generar insights valiosos, sino también en su potencial para adaptarse a las necesidades cambiantes del mercado laboral y las expectativas de los estudiantes.

El impacto de estas tecnologías en la empleabilidad de los graduados es significativo, ofreciendo no solo una mejor alineación con las demandas del mercado laboral sino también promoviendo una mayor integración entre la educación superior y el sector empresarial. A medida que las universidades adoptan estas herramientas, se posicionan mejor para cumplir con su misión de preparar a los estudiantes no solo para los trabajos de hoy, sino también para las carreras del futuro.

En conclusión, la integración de la analítica de datos en las estrategias para mejorar la empleabilidad de los graduados representa un cambio paradigmático en la educación superior. Los modelos de clustering y los sistemas de recomendación no solo facilitan una gestión más efectiva de las relaciones con el sector empresarial y una orientación profesional más personalizada, sino que también señalan el camino hacia una educación que es simultáneamente más adaptativa y alineada con las realidades del mercado laboral contemporáneo. Este enfoque, centrado en la aplicación práctica de la analítica de datos, no solo mejora la empleabilidad de los graduados, sino que también refuerza el papel de las instituciones educativas como facilitadores clave en la transición de la educación al empleo.

Bibliografía

1. Abelha, M., et al. (2020). Mismatch Between University Graduates' Competencies and Employers' Needs. *Journal of Higher Education and Workforce Development*, 12(3), 45-59.

2. Ashari, I. F., Nugroho, E. D., Baraku, R., Yanda, I. N., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. *Journal of Applied Informatics and Computing*, 7(1). <https://dx.doi.org/10.30871/jaic.v7i1.4947>
3. Balabanović, M., & Shoham, Y. (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66-72.
4. Blinov, V., Sergeev, I. S., & Rodichev, N. F. (2022). Microlearning in Business Education: A Promising Direction.
5. Bowen, W. (1997). Experimentos de campo y el arte de los modelos. *CIP-Circular*, 22(3), 23.
6. Brusil, C. (2024). Modelo Cluster. Recuperado de https://github.com/Carlosmagnobrusil/CarlosBrusil_projects/blob/main/Sistema_de_recomendacion.ipynb
7. Brusil, C. (2024). Sistema de recomendación. Recuperado de https://github.com/Carlosmagnobrusil/CarlosBrusil_projects/blob/main/Tratamiento_de_datos_Modelo_Cluster.ipynb
8. Brusil, C. (2024). Tratamiento de datos Modelo Cluster. Recuperado de https://github.com/Carlosmagnobrusil/CarlosBrusil_projects/blob/main/Modelo_Cluster.ipynb
9. Cadet, N., & Griffiths, T. (2023). Embedding Employability in the Social Sciences: A Case Study from a Post-92 Higher Education Institution in the UK. *Journal of Perspectives in Applied Academic Practice*, 11(2). <https://dx.doi.org/10.56433/jpaap.v11i2.551>
10. Cabrera, J. J. D., Sison, A. M., & Medina, R. P. (2019). Enhanced initialization method for the K-means algorithm using radius and 360 degrees for centroid positions. *Proceedings of the ACM Symposium on Applied Computing*, 335-342.
11. Cantador, I., Brusilovsky, P., & Kuflik, T. (2018). Second workshop on information heterogeneity and fusion in recommender systems (HetRec

- 2011). Proceedings of the 5th ACM conference on Recommender systems, 387-388.
12. Cheowsuwan, T. (2016). Factors Affecting Fraudulent Financial Reporting in Companies Implementing ASEAN Corporate Governance Scorecard. *Journal of Financial Reporting*, 11(4), 58-76.
 13. Davidson, I. (2002). Understanding K-means non-hierarchical clustering. Computer Science Department of State University of New York (SUNY), Albany.
 14. Davis, A., & Thompson, R. (2023). Enhancing Recommendation Systems through Cross-Validation Techniques. *Journal of Data Science*, 11(2), 134-145.
 15. Draissi, Z., Rong, Y., & Suliman, M. A. E. (2023). Skill Mismatch in Urban Workforce of Morocco. *Journal of Workforce Development*.
 16. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96).
 17. Fidan, M. (2023). Effects of a Microlearning-Supported Flipped Classroom on Pre-Service Teachers' Learning Performance, Motivation, and Engagement. *Education and Information Technologies*. <https://dx.doi.org/10.1007/s10639-023-11639-2>
 18. Flornoy-Guédon, A., Fonzo-Christe, C., Meier, E., Gazengel-Marchand, M., François, O., Gschwind, L., & Bonnabry, P. (2023). Development and Evaluation of a Blended Learning Training Program for Pharmacy Technicians.
 19. Guerra, C. W., Cabrera, A., & Fernández, L. (2003). Criterios para la selección de modelos estadísticos en la investigación científica. *Revista Cubana de Ciencia Agrícola*, 37(1), 3-10.
 20. Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques*. Elsevier.

21. Hanshaw, G., & Hanson, J. (2019). Impact of Microlearning and Social Learning on Teacher Professional Development. *International Journal of Learning, Development and Educational Research*, 9(1). <https://dx.doi.org/10.5296/IJLD.V9I1.13713>
22. Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5-53.
23. Hu, S., Jiang, L., & Chen, L. (2022). Help or Hindrance? Perceived Employability and its Cross-Lagged Relations with Job Insecurity and Job-Related Well-Being.
24. Hu, S., Jiang, L., & Chen, L. (2022). The Cross-Lagged Relations Between Perceived Employability and Job Insecurity: A Three-Wave Study. *European Journal of Work and Organizational Psychology*. <https://dx.doi.org/10.1080/1359432X.2022.2050219>
25. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>
26. Javed, U., Shaukat, K., Hameed, I., Iqbal, F., Alam, T. M., & Luo, S. (2021). Context-aware and content-based recommender systems for news recommendation. *International Journal of Emerging Technologies in Learning*, 16(03). <https://doi.org/10.3991/IJET.V16I03.18851>
27. Johnson, H., & Lee, A. (2022). Benchmarking the Precision of Content-Based Recommendation Systems. *International Journal of Computer Science*, 19(4), 88-97.
28. Kaggle; Shubham Chambhare. (s.f.). All Major Fields Courses in Udemy - 44K Courses. Recuperado de <https://www.kaggle.com/datasets/shubhamchambhare/all-major-fields-courses-in-udemy-44k-courses>
29. Kankam, G. (2022). Are Relation Strategies Useful for Ghanaian Public Universities?
30. Kaminski, L., Lavadenz, M., & Armas, E. G. (s.f.). Impact of the Bilingual Teacher Professional Development Program in California. Center for Equity

for English Learners, Loyola Marymount University.
<https://dx.doi.org/10.15365/ceel.policy.13>

31. Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. John Wiley & Sons.
32. Khuziakhmetov, R., & Romashkina, G. (2022). Students' Employability Skills Development Strategies.
33. Lázár, T., Farkas, J., Hajdu, Z., Tar, I., Czeller, M., & D'Arrigo, M. (2023). Intercultural Competencies of University Students and Their Relation to Employability. *Central European Journal of Educational Research*, 5(2).
<https://dx.doi.org/10.37441/cejer/2023/5/2/13222>
34. Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50, 57-70.
35. Leijon, M., Algotson, A., Bernhardsson, S., Ekholm, D., & coautores. (2024). Generation Pep-Study protocol for an intersectoral community-wide physical activity and healthy eating habits initiative for children and young people in Sweden. *Frontiers in Public Health*. Disponible en:
<https://www.frontiersin.org/articles/10.3389/fpubh.2024.1299099/full>
36. Liao, M., & Sundar, S. S. (2021). E-commerce consumers' response to product recommendations: Collaborative filtering, content-based filtering, and the persuasiveness of social proof. *Journal of Advertising*, 50(1), 84-99.
<https://doi.org/10.1080/00913367.2021.1887013>
37. Lopez, S., Martinez, J., & Garcia, P. (2024). Addressing Bias and Diversity in Educational Recommendation Systems. *Educational Technology Research and Development*, 22(3), 205-220.
38. Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook* (pp. 73-105). Springer, Boston, MA.
39. Lukasik, S., Kowalski, P. A., Charytanowicz, M., & Kulczycki, P. (2016). Clustering using flower pollination algorithm and Calinski-Harabasz index.

- IEEE Congress on Evolutionary Computation (CEC).
<https://dx.doi.org/10.1109/CEC.2016.7744132>
40. Mainga, W., Daniel, R., & Alamil, L. R. (2022). Perceptions of Employability Skills Among Undergraduate Business Students and Industry Representatives: A Qualitative Study.
41. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
42. Maulik, U., & Bandyopadhyay, S. (2002). Performance Evaluation of Some Clustering Algorithms and Validity Indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12).
<https://dx.doi.org/10.1109/TPAMI.2002.1114856>
43. McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 445, 51-56.
<https://conference.scipy.org/proceedings/scipy2010/mckinney.html>
44. Mutula, P. K. O., & Engairo, P. (2024). Effect of Risk Identification on Project Implementation among Faith Based Construction Projects in the Diocese Of Ngong. *Researchbridgepublisher.com*. Disponible en:
<https://researchbridgepublisher.com/index.php/ijsshr/article/view/67>
45. Niknam, T., & Amiri, B. (2010). An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. *Applied Soft Computing*, 10(1), 183-197. Disponible en:
<https://www.sciencedirect.com/science/article/abs/pii/S1568494609000854>
46. Oliphant, T. E. (2006). A guide to NumPy. Trelgol Publishing.
47. Pazzani, M., & Billsus, D. (n.d.). Content-based recommendation systems. In *The Adaptive Web*. https://doi.org/10.1007/978-3-540-72079-9_10
48. Pulis, M., & Bajada, J. (2021). Siamese neural networks for music recommendation. *ACM Digital Library*.
<https://doi.org/10.1145/3460231.3478847>
49. Quintero-Beltrán, L. C., & Osorio-Morales, L. M. (2018). Balanced Scorecard como herramienta para empresas en estado de crisis. *Revista CEA*.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3519380

50. Rahmah, N., & Sitanggang, I. S. (2016). Determination of optimal epsilon (eps) value on dbscan algorithm to clustering data on peatland hotspots in Sumatra. In IOP Conference Series: Earth and Environmental Science (Vol. 31, No. 1, p. 012012). IOP Publishing.
51. Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to Recommender Systems Handbook. Springer.
52. Romanuke, V. (2023). Using rectangular and hexagonal clusters for initial centroid calculation in the K-means algorithm. *Applied Computer Systems*, 27(1), 1-6.
53. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.
54. Sachitra, K., & Kumara, U. A. (2019). The Employability of External Degree Holders in Management: A Case Study Based on the University of Sri Jayewardenepura, Sri Lanka. *Asian Journal of Education and Social Studies*, 4(1), 1-12. <https://dx.doi.org/10.9734/AJESS/2019/V4I130112>
55. Salazar, J. G. N., & Vergara-Romero, A. (2022). Strategies for optimizing administrative management in a construction company using the balanced scorecard. *Res non verba*. <https://revistas.ecotec.edu.ec/index.php/rnv/article/view/623>
56. Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 19.
57. Senarath, S. A. C. L., & Patabendige, S. (2015). Balanced Scorecard Techniques for Corporate Plan Preparation in Sri Lankan Universities. *International Journal of University Corporate Relations*, 8(2), 113-127.
58. Shahid, S., & Ayyaz, I. (2023). Employer Branding Perception and the Role of Social Media and Prior Work Experience on the Job Application Intention.
59. Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender Systems Handbook* (pp. 257-297). Springer.

60. Shoss, M. K., Su, S., Schlotzhauer, A. E., & Carusone, N. (2022). Working Hard or Hardly Working: An Examination of the Preservation Motivation Perspective on Job Insecurity.
61. Soim, S., Haryanti, N., & Nurohmah, N. (2022). Public Relations Management in College: A Case Study.
62. Son, J., & Kim, S. (2017). A content-based recommendation system using a multiattribute network. *Expert Systems with Applications*, 89, 1-9. <https://doi.org/10.1016/j.eswa.2017.08.008>
63. Sozmen, E. (2022). A Perspective on the Pros and Cons of Microlearning in Health Education.
64. Superintendencia de Compañías. (2023). *bi_ranking.csv*. Recuperado de <https://appscvsmovil.supercias.gob.ec/ranking/reporte.html>
65. Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to data mining*. Pearson Education.
66. Thorat, S., Ashwini, G., & Seema, M. (2023). Recommendation models: Collaborative filtering and content-based filtering. *IEEE Xplore*. <https://doi.org/10.1109/ICSSIT55814.2023.10061072>
67. Vega Falcón, V., & Jácome, D. K. L. (2020). The Balanced Scorecard as an Organizational Management Tool. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3521792
68. Volkova, N. (2023). Professional Development in Education: An Overview. *Journal of Education and Training*. <https://dx.doi.org/10.54891/2786-698x-2023-1-6>
69. Waskom, M. L., et al. (2020). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 5(53), 2307. <https://doi.org/10.21105/joss.02307>
70. Wibowo, N. L., Soeleman, M., & Fanani, A. Z. (2023). Modification of the K-means algorithm using the Antlion Optimizer for improved initial centroid determination. *Journal of Theoretical and Applied Information Technology*, 7(4), 4997-5008.
71. Wong, L. M., Sutherland, T., Perry, E., Tran, V., & coautores. (2024). Fluorine-18-labelled Prostate-specific Membrane Antigen Positron Emission

- Tomography/Computed Tomography or Magnetic Resonance Imaging to. European Urology. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2588931124000269>
72. Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on Neural Networks, 16(3), 645-678.
73. Zamani, Z., Tamartash, R., Heydari, Q. A., & coautores. (2024). Investigating the environmental factors affecting the quantitative and qualitative performance of the medicinal plant *Stachys lavandulifolia* Vahl. Journal of Environmental Science. Disponible en: https://www.jess.ir/article_189304.html?lang=en

Anexos

Anexo 1. Balance Score Card para el área de relaciones Externas.

Anexo 2. Base de datos homologada

| # | Variable | Naturaleza Principal | Subdivisión |
|----------|--------------------------|-----------------------------|--------------------|
| 0 | unnamed:_0 | Cuantitativa | Discreta |
| 1 | organization_name | Cualitativa | Nominal |
| 2 | primary_contact | Cualitativa | Nominal |
| 3 | contact:_employer(count) | Cuantitativa | Discreta |
| 4 | email | Cualitativa | Nominal |
| 5 | phone | Cualitativa | Nominal |
| 6 | created | Cuantitativa | Discreta |
| 7 | matches | Cualitativa | Nominal |

| | | | |
|----|----------------------------------|--------------|----------|
| 8 | matched_name | Cualitativa | Nominal |
| 9 | similarity_score | Cuantitativa | Continua |
| 10 | match_index | Cuantitativa | Discreta |
| 11 | no._fila | Cuantitativa | Discreta |
| 12 | expediente_x | Cuantitativa | Discreta |
| 13 | ruc | Cuantitativa | Discreta |
| 14 | nombre_x | Cualitativa | Nominal |
| 15 | situación_legal | Cualitativa | Nominal |
| 16 | fecha_constitucion | Cualitativa | Nominal |
| 17 | tipo | Cualitativa | Nominal |
| 18 | país | Cualitativa | Nominal |
| 19 | región | Cualitativa | Nominal |
| 20 | provincia_x | Cualitativa | Nominal |
| 21 | cantón | Cualitativa | Nominal |
| 22 | ciudad_x | Cualitativa | Nominal |
| 23 | calle | Cualitativa | Nominal |
| 24 | número | Cualitativa | Nominal |
| 25 | intersección | Cualitativa | Nominal |
| 26 | teléfono | Cualitativa | Nominal |
| 27 | representante | Cualitativa | Nominal |
| 28 | cargo | Cualitativa | Nominal |
| 29 | capital_suscrito | Cualitativa | Nominal |
| 30 | ciiu_nivel_1 | Cualitativa | Nominal |
| 31 | ciiu_nivel_6 | Cualitativa | Nominal |
| 32 | posición | Cuantitativa | Discreta |
| 33 | año | Cuantitativa | Discreta |
| 34 | pocisión_ | Cuantitativa | Discreta |
| 35 | año.1 | Cuantitativa | Discreta |
| 36 | expediente_y | Cuantitativa | Discreta |
| 37 | nombre_y | Cualitativa | Nominal |
| 38 | tipo_compañía | Cualitativa | Nominal |
| 39 | actividad_económica | Cualitativa | Nominal |
| 40 | región_ | Cualitativa | Nominal |
| 41 | provincia_y | Cualitativa | Nominal |
| 42 | ciudad_y | Cualitativa | Nominal |
| 43 | tamaño | Cualitativa | Nominal |
| 44 | sector | Cualitativa | Nominal |
| 45 | cant._empleados | Cuantitativa | Discreta |
| 46 | activo_2022 | Cuantitativa | Continua |
| 47 | patrimonio_2022 | Cuantitativa | Continua |
| 48 | ingreso_por_ventas_2022 | Cuantitativa | Continua |
| 49 | utilidad_antes_del_impuesto_2022 | Cuantitativa | Continua |
| 50 | utilidad_del_ejercicio_2022 | Cuantitativa | Continua |
| 51 | utilidad_neta_2022 | Cuantitativa | Continua |
| 52 | ir_causado_2022 | Cuantitativa | Continua |

| | | | |
|----|--------------------|--------------|----------|
| 53 | ingreso_total_2022 | Cuantitativa | Continua |
|----|--------------------|--------------|----------|

Anexo 3. Lista de la clasificación sectorial CIIU

| CODIGO | DESCRIPCION | NIVEL |
|---------------|--|--------------|
| A | AGRICULTURA, GANADERÍA, SILVICULTURA Y PESCA. | 1 |
| B | EXPLOTACIÓN DE MINAS Y CANTERAS. | 1 |
| C | INDUSTRIAS MANUFACTURERAS. | 1 |
| D | SUMINISTRO DE ELECTRICIDAD, GAS, VAPOR Y AIRE ACONDICIONADO. | 1 |

| | | |
|---|---|---|
| E | DISTRIBUCIÓN DE AGUA; ALCANTARILLADO, GESTIÓN DE DESECHOS Y ACTIVIDADES DE SANEAMIENTO. | 1 |
| F | CONSTRUCCIÓN. | 1 |
| G | COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACIÓN DE VEHÍCULOS AUTOMOTORES Y MOTOCICLETAS. | 1 |
| H | TRANSPORTE Y ALMACENAMIENTO. | 1 |
| I | ACTIVIDADES DE ALOJAMIENTO Y DE SERVICIO DE COMIDAS. | 1 |
| J | INFORMACIÓN Y COMUNICACIÓN. | 1 |
| K | ACTIVIDADES FINANCIERAS Y DE SEGUROS. | 1 |
| L | ACTIVIDADES INMOBILIARIAS. | 1 |
| M | ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICAS. | 1 |
| N | ACTIVIDADES DE SERVICIOS ADMINISTRATIVOS Y DE APOYO. | 1 |
| O | ADMINISTRACIÓN PÚBLICA Y DEFENSA; PLANES DE SEGURIDAD SOCIAL DE AFILIACIÓN OBLIGATORIA. | 1 |
| P | ENSEÑANZA. | 1 |
| Q | ACTIVIDADES DE ATENCIÓN DE LA SALUD HUMANA Y DE ASISTENCIA SOCIAL. | 1 |
| R | ARTES, ENTRETENIMIENTO Y RECREACIÓN. | 1 |
| S | OTRAS ACTIVIDADES DE SERVICIOS. | 1 |
| T | ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES; ACTIVIDADES NO DIFERENCIADAS DE LOS HOGARES COMO PRODUCTORES DE BIENES Y SERVICIOS PARA USO PROPIO. | 1 |
| U | ACTIVIDADES DE ORGANIZACIONES Y ÓRGANOS EXTRATERRITORIALES. | 1 |

