



PROYECTO FIN DE GRADO: MODELO PREDICTIVO DE RENUNCIA VOLUNTARIA EN LA EMPRESA

Facultad de Ciencias Económicas y empresariales
(ICADE)

Autor: Cristina Hernández Ayúcar

Tutor: María José Martín Rodrigo

RESUMEN

El presente trabajo de fin de grado tiene como propósito principal la presentación y validación de un modelo predictivo de renuncia voluntaria de empleados en las empresas.

Este estudio destaca la importancia crucial de integrar Business Analytics para optimizar procesos en recursos humanos a través de "People Analytics". El objetivo principal es reducir la alta rotación de personal identificando los factores clave que motivan a los empleados a dejar sus puestos. Los resultados obtenidos permitirán desarrollar estrategias de retención más efectivas, mejorando la gestión de recursos humanos y reduciendo los costos asociados a la rotación de personal.

Con el fin de mejorar la satisfacción en el ambiente laboral se ha desarrollado un modelo predictivo de renuncia voluntaria de empleados.

Palabras clave: *Renuncia voluntaria, modelo predictivo, recursos humanos, algoritmo, satisfacción laboral, análisis de datos.*

ABSTRACT

The main purpose of this final degree project is to present and validate a predictive model for voluntary employee turnover in companies.

This study highlights the crucial importance of integrating Business Analytics to optimize human resources processes through "People Analytics."

The primary objective is to reduce high employee turnover by identifying the key factors that motivate employees to leave their positions. The results obtained will allow us to develop more effective retention strategies, improving human resources management and reducing the costs associated with employee turnover.

In order to enhance job satisfaction, we have developed an innovative predictive model for voluntary employee turnover.

Keywords: Voluntary resignation, predictive model, human resources, algorithm, job satisfaction, data analysis.

Contents

1 INTRODUCCIÓN	4
1.2 Contexto y motivación	4
1.2 Planteamiento de la necesidad a cubrir	6
1.3 Objetivo de la investigación	9
1.4 Metodología	11
2 MARCO TEÓRICO	13
2.1 Definición de RRHH, objetivos y funciones	13
2.2 La digitalización de los RRHH en España	16
2.3 Modelos predictivos en RRHH y su impacto económico en las empresas.....	19
3 INVESTIGACIÓN MEDIANTE ANÁLISIS DE DATOS.....	22
3.1 Preparación del conjunto de datos	22
3.2 Análisis descriptivo.....	28
3.3 Modelo explicativo.....	33
3.4 Modelos predictivo	37
4 ANÁLISIS DE RESULTADOS.....	49
4.1 Interpretación de resultados.....	49
4.5 Visualizaciones	52
5 DISCUSIÓN.....	65
5.1 Literatura existente	65
5.2 Implicaciones prácticas del modelo	69
6 CONCLUSIONES	73
7 DECLARACIÓN RESPECTO AL USO DE CHAT GPT U OTRAS HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA.	75
8 BIBLIOGRAFÍA.....	77

1 INTRODUCCIÓN

1.2 Contexto y motivación

El objetivo principal de este trabajo de fin de grado es lograr una integración efectiva entre el ámbito empresarial y la disciplina de Business Analytics. Este estudio se propone explorar cómo las herramientas y técnicas de análisis de datos pueden aplicarse para resolver problemas empresariales, optimizar procesos y tomar decisiones estratégicas más informadas. A través de una investigación exhaustiva y la aplicación práctica de métodos analíticos, se busca demostrar la relevancia y el impacto positivo que puede tener la analítica de negocios en el entorno corporativo (Martín,2023).

Es indudable que el análisis de datos ha adquirido una relevancia creciente en el ámbito empresarial. Esta tendencia se ha visto impulsada aún más por el reciente auge de la "inteligencia generativa", una innovadora rama de la inteligencia artificial que está transformando diversos sectores. Ante este panorama, las empresas se enfrentan a la necesidad imperiosa de adaptarse a estos avances tecnológicos. Para mantener y mejorar su posición competitiva en el mercado, resulta crucial que las organizaciones incorporen gradualmente estas nuevas herramientas y metodologías en sus procesos operativos y estratégicos (ISDI, 2023).

En consecuencia, las herramientas asociadas al ámbito de Business Analytics están siendo cada vez más implementadas en todos los sectores de la empresa con el propósito de optimizar la eficiencia operativa y aumentar la productividad. Estas tecnologías permiten a las organizaciones analizar grandes volúmenes de datos, identificar patrones y tendencias, y tomar decisiones informadas que mejoran significativamente los resultados en diversas áreas (EAE Business School, 2023).

Este trabajo de investigación se enfocará específicamente en el área de recursos humanos, un campo donde el potencial de mejora mediante el análisis de datos es aun considerablemente amplio. A pesar de los avances ya logrados, los recursos humanos pueden beneficiarse enormemente de la integración de Business Analytics para gestionar

el talento, predecir necesidades futuras, mejorar la satisfacción y el rendimiento de los empleados, y optimizar los procesos de selección y retención de personal.

El concepto emergente de "People Analytics" se refiere a la aplicación de técnicas de análisis de datos cualitativos y cuantitativos relacionados con los empleados para mejorar el rendimiento organizacional. Este enfoque permite a las empresas tomar decisiones estratégicas basadas en datos concretos en lugar de la intuición (Pérez, 2023).

La analítica de recursos humanos (RR.HH.) implica recopilar, procesar y analizar diversos datos de empleados, como niveles de productividad, satisfacción laboral e historial de desempeño. Esto proporciona una comprensión más precisa de la fuerza laboral (Peeters et al., 2020).

Por ejemplo, People Analytics puede identificar patrones que indiquen la intención de los empleados de abandonar la empresa, permitiendo intervenciones proactivas para retener talento clave. Además, revela el nivel de compromiso y expectativas de desarrollo profesional de los empleados, facilitando la creación de estrategias de desarrollo y planes de carrera personalizados (Tursunbayeva et al., 2018).

Asimismo, People Analytics evalúa el impacto de políticas de RR.HH. como escalas salariales, beneficios y programas de formación. Analizando cómo estas políticas afectan la retención y satisfacción de los empleados, las empresas pueden ajustar sus enfoques para minimizar la fuga de talento y maximizar el rendimiento y bienestar de su fuerza laboral (Wolterskluwer, 2023).

En estudios previos, como el de Fitz-enz, se ha demostrado que People Analytics puede ser utilizado para identificar empleados con alto potencial y desarrollar estrategias de retención efectivas. Fitz-enz señala que, mediante el análisis de datos históricos de desempeño y satisfacción laboral, las organizaciones pueden predecir qué empleados tienen mayores probabilidades de renunciar y tomar medidas preventivas (Fitz-Enz & Mattox, 2014).

En resumen, People Analytics transforma la manera en que las empresas gestionan a su personal, proporcionando una base sólida de datos para la toma de decisiones estratégicas y mejorando la eficiencia y efectividad de las políticas de recursos humanos.

1.2 Planteamiento de la necesidad a cubrir

La fuga de talento y la elevada rotación de personal son problemas frecuentes en el ámbito empresarial actual. Los recursos humanos ocupan una posición estratégica en cualquier organización, ya que su función principal es liderar, coordinar, mantener, mejorar y desarrollar las metas establecidas por la empresa a través de una gestión eficaz del capital humano. Sin embargo, resulta extremadamente complicado lograr estos objetivos cuando se trabaja con una plantilla laboral en constante cambio.

Un índice de rotación elevado generalmente no es compatible con la estabilidad laboral necesaria para el éxito organizacional. En este contexto, cualquier esfuerzo que se pretenda implementar para mejorar la organización se enfrenta a grandes desafíos, ya que es como cultivar en un terreno infértil. Es difícil consolidar programas, filosofías o sistemas empresariales cuando la base sobre la que se trabaja es inestable. Es cierto que toda empresa experimenta cierto grado de rotación, lo cual puede tener aspectos positivos, como la renovación de ideas y la incorporación de nuevas perspectivas. No obstante, los costos asociados a la alta rotación son significativos, y las empresas buscan minimizarla para mantener un entorno laboral más estable y productivo. Detectar la inestabilidad laboral generada por la rotación elevada es el primer paso, pero es igualmente crucial investigar las causas subyacentes de este fenómeno.

Los altos niveles de rotación suelen estar asociados a diversos factores, entre los que destacan problemas de liderazgo, la falta de trabajo en equipo, la ausencia de motivación y la insatisfacción laboral. Un liderazgo deficiente puede provocar una falta de dirección y apoyo, lo que disminuye la motivación y el compromiso de los empleados. La falta de cohesión en los equipos de trabajo impide que los empleados sientan que forman parte de una unidad, lo que puede aumentar la sensación de aislamiento y descontento.

Además, la insatisfacción laboral puede ser causada por múltiples factores, como la falta de oportunidades de desarrollo profesional, condiciones laborales inadecuadas,

falta de reconocimiento y recompensas, y una cultura organizacional que no fomente el bienestar y el crecimiento de sus empleados. Identificar y abordar estas causas es esencial para reducir la rotación de personal y crear un entorno de trabajo más saludable y sostenible (López,2011).

Se ha recopilado información de varios estudios con el objetivo de identificar los costos principales que enfrenta una empresa ante altos niveles de rotación de personal. A continuación, se detallan estos costos y su impacto en la organización:

- **Costos de Reclutamiento y Selección**

Uno de los costos más significativos es el tiempo y los recursos invertidos en el proceso de selección de nuevos empleados. Esto incluye entrevistas, evaluaciones y la publicidad de las vacantes, tanto en línea como en medios físicos. La empresa debe invertir en plataformas de contratación externas para llegar a un mayor número de candidatos, lo que incrementa significativamente los gastos (Estrada,2024).

- **Gastos de Capacitación e Inducción**

Al incorporar nuevos empleados, la empresa incurre en gastos relacionados con programas de formación y el tiempo que otros miembros del equipo deben dedicar a capacitar y preparar a los recién llegados. Este proceso de inducción es esencial para que los nuevos empleados se integren adecuadamente en la plantilla y adquieran las habilidades necesarias para su desempeño (Castillo Pizo,2021).

- **Pérdida de Productividad**

Es importante considerar la pérdida de productividad durante el periodo de adaptación de los nuevos empleados. Este tiempo es necesario para que alcancen el nivel de eficiencia de sus predecesores, lo que puede reducir la operatividad del equipo en general. Además, los nuevos trabajadores son más propensos a cometer errores durante su integración, lo que puede generar retrasos y disminuir la eficacia operativa (Huanuco Naupay & Santillan Carrion, 2023).

- **Costos Administrativos**

La salida y la incorporación de nuevos empleados conllevan una serie de costos administrativos. Esto incluye el papeleo y los trámites necesarios, así como la actualización de los sistemas de recursos humanos y la modificación de registros. Estos procesos administrativos son inevitables y representan un gasto adicional para la empresa (Quintero Mejía & Ordoñez Barajas, 2023)

- **Impacto en la Moral y el Clima Laboral**

La alta rotación de personal puede afectar negativamente el clima laboral. Los empleados que permanecen en la empresa pueden experimentar desmotivación y estrés, lo que deteriora el trabajo en equipo y reduce la cohesión y colaboración entre los miembros del equipo. Para contrarrestar esto, la empresa podría verse obligada a incrementar los costos de retención mediante bonificaciones y otros incentivos, lo que aumenta aún más los gastos.

- **Pérdida de Conocimiento y Experiencia**

La fuga de talento implica la pérdida de empleados con conocimientos críticos y habilidades valiosas para la organización. Esto es comparable a perder una cosecha después de haber dedicado tiempo y recursos a cultivarla y hacerla crecer. La pérdida de estos recursos humanos puede ser devastadora para la continuidad y el éxito de la empresa (Guerrero Velástegui & Parra Real ,2023).

- **Impacto en la Imagen y la Reputación**

Finalmente, un alto nivel de rotación puede dañar la reputación e imagen de la empresa en el mercado laboral. Esto puede dificultar la atracción de nuevo talento en el futuro, obligando a la empresa a ofrecer salarios más altos o beneficios adicionales para atraer a nuevos empleados. Esta percepción negativa puede tener efectos duraderos y perjudiciales para la empresa en el largo plazo (Estrada,2024).

En resumen, la estabilidad y el desarrollo de cualquier empresa dependen en gran medida de una gestión eficaz de los recursos humanos. Asimismo, la alta rotación de personal genera una serie de costos directos e indirectos que afectan la operatividad y la estabilidad de una empresa. Es crucial implementar estrategias de retención efectivas para

minimizar estos costos y mantener un entorno laboral saludable y productivo (Grisales & Viviana, 2020)

1.3 Objetivo de la investigación

Una vez identificados todos los problemas que los altos niveles de rotación pueden acarrear a la empresa, resulta evidente el significativo impacto que un modelo predictivo de renuncia voluntaria por parte de los empleados podría tener a nivel empresarial. Por ello, el objetivo principal de esta investigación es desarrollar un modelo predictivo de clasificación binaria utilizando la técnica de regresión logística.

El propósito de este modelo es doble: por un lado, identificar las variables más influyentes que afectan la decisión de los empleados de renunciar voluntariamente a su puesto de trabajo, y, por otro lado, predecir con precisión si un empleado abandonará su puesto de trabajo de manera voluntaria.

Para lograr estos objetivos, se analizarán diversos factores que podrían influir en la decisión de renuncia, tales como el nivel de satisfacción laboral, las oportunidades de crecimiento profesional, el equilibrio entre la vida laboral y personal, la cultura organizacional, y otros indicadores relevantes. La identificación de estas variables permitirá no solo entender mejor las razones detrás de la rotación voluntaria, sino también implementar estrategias efectivas para retener el talento dentro de la organización.

En resumen, este estudio no solo aspira a predecir la renuncia voluntaria de los empleados, sino también a proporcionar información valiosa que permita a la empresa tomar decisiones informadas y diseñar políticas que fomenten la retención y satisfacción de los empleados, contribuyendo así al fortalecimiento y estabilidad de la empresa a largo plazo.

En la actualidad existen una gran cantidad de teorías sobre por qué los empleados deciden dejar sus trabajos. No obstante, el objetivo es probar con datos aquello que sustentan estas teorías. A continuación, se mencionan algunas teorías principales, a fin de identificar las causas que ponen de manifiesto como desencadenante en la renuncia voluntaria en puestos de trabajo. Para posteriormente, ver si la investigación y los resultados obtenidos apoyan esas hipótesis ya establecidas.

Entre ellas destacan la teoría de la equidad y la teoría de las expectativas.

La teoría de la equidad fue desarrollada por John Stacey Adams en 1963 y tiene mucha relevancia dentro del ámbito de la gestión de recursos humanos. Esta teoría permite comprender y aprender a gestionar la motivación de los empleados centrándose en la percepción de justicia y equidad dentro de una compañía (Puyol-Cortez, 2021). Sostiene el concepto de que los trabajadores persiguen el equilibrio entre lo que aportan y lo que reciben. Entre los principales “inputs” destacan la experiencia, dedicación, esfuerzo, competencias o habilidad. Y el reconocimiento, los beneficios para el empleado, el salario o promociones son reconocidos como posibles “outputs”. Para conseguir mantener a los empleados motivados y satisfechos, ellos mismos deben sentir que la empresa trata de manera equitativa la relación entre “inputs” y “outputs” de todos los empleados. Es decir, deben sentir que sus esfuerzos son debidamente recompensados y no percibir desigualdad en comparación a lo que reciben otros empleados. Por eso resulta de vital importancia la transparencia en la gestión de recompensas y fomentar un entorno de trabajo justo y equilibrado (Fernández Palma, 2009)

En segundo lugar, se destaca la teoría de las expectativas desarrollada en 1964 por Victor H. Vroom. Esta teoría trata de identificar qué incita a los trabajadores a estar motivados a alcanzar ciertos objetivos. La teoría sostiene que los trabajadores se ven influenciados por tres factores: expectativa, instrumentalidad y valencia (Marrugo & Pérez Meléndez, 2012).

El primer factor, la expectativa, hace referencia a tener la percepción de que el esfuerzo bien hecho traerá una recompensa satisfactoria. La instrumentalidad establece que un buen desempeño te hará lograr una buena recompensa. Y la valencia hace referencia al valor que el empleado asigna a la posible recompensa. Según la teoría de las expectativas la fórmula de la motivación del empleado se explica por como interactúan estos tres componentes. Si queremos tener empleados altamente motivados todos estos factores deben ser positivos y significativos, en caso contrario la motivación disminuirá (López Chánez et al. 2007).

Tras desarrollar ambas teorías, se identifican posibles causas que pueden ser desencadenantes de la renuncia voluntaria en las empresas. Entre ellas destacan el desequilibrio entre lo que aportan los empleados y lo que reciben, y la falta de motivación.

Ahora se realiza un trabajo de investigación a fin de comprobar si lo que establecen estas y otras teorías que mencionaremos más adelante es cierto. Y en caso de que lo sea diseñar estrategias en el ámbito de recursos humanos que den solución a los problemas identificados.

1.4 Metodología

Tras haber identificado en las empresas la necesidad de transformar la manera en la que gestionan su personal, implementando “People Analytics”, se establecen una serie de pasos a seguir.

Primeramente, se lleva a cabo una investigación previa en el campo de los recursos humanos y su posible relación con el “machine learning”. Se considera necesario establecer bien las bases de nuestra investigación y saber cual es la situación actual de la digitalización, análisis de datos y modelos predictivos en el ámbito de gestión de capital humano, para poder después desarrollar nuestro modelo.

Seguidamente, se iniciará el proceso de investigación y análisis de datos. Se comenzará con la preparación del “dataset”, identificando todas las variables que posee, examinando los valores de estas y eliminando aquellas que sean inservibles. Después, se realizará un análisis descriptivo a fin de obtener un resumen de ciertos estadísticos de cada variable, ver las relaciones que se establecen entre ellas e identificar patrones. A continuación, se realizará un modelo explicativo que permitirá identificar aquellas variables que tienen una mayor influencia en los empleados a la hora de tomar la decisión de abandonar su puesto de trabajo. Más adelante, llegara el momento de implementar los modelos predictivos.

-Se realizarán dos modelos utilizando el algoritmo de regresión logística (LOGIT) que permita llevar a cabo una clasificación binaria, obteniendo así la probabilidad de que un empleado abandone la empresa. Y posteriormente, a fin de mejorar el rendimiento, se generará un tercer modelo utilizando el algoritmo “RandomForest”.

Una vez terminada la fase de investigación, se interpretarán los resultados obtenidos y se realizarán una serie de visualizaciones a fin de clarificar las conclusiones generadas por el modelo.

Adicionalmente, se compararán las conclusiones obtenidas con la literatura ya existente, para dar veracidad y probar la congruencia de nuestros hallazgos.

Por último, se redactarán una serie de conclusiones a fin de recopilar toda la información generada a lo largo de la investigación, a modo de cierre del presente trabajo.

2 MARCO TEÓRICO

2.1 Definición de RRHH, objetivos y funciones

El área de Recursos Humanos (RRHH) en una empresa se dedica a la supervisión y manejo de los colaboradores. Originalmente, esta disciplina se enfocaba en tareas administrativas relacionadas con el personal, pero con el tiempo ha evolucionado para incorporar responsabilidades alineadas con la visión y estrategia del equipo humano. Se reconoce que la Revolución Industrial marcó el inicio de una estructuración laboral más efectiva. A finales del siglo XIX, se definieron los fundamentos de la organización sistemática del trabajo, dando origen a los esquemas organizativos que conocemos hoy día.

El concepto de Recursos Humanos se origina como una categoría que abarca a todos los individuos asociados con una entidad, ya sea una empresa, organización o sector específico. Se trata de la totalidad del personal y su relación laboral con la entidad. En el contexto de la gestión empresarial, RRHH se refiere a la administración integral que aborda todos los procesos relacionados con los trabajadores, desde el momento previo a su incorporación hasta el final de su relación laboral.

La gestión de RRHH incluye una variedad de procesos esenciales como la selección, formación, evaluación y promoción del personal. Además, esta administración va más allá de las tareas burocráticas, englobando también la promoción del compromiso y la motivación de los trabajadores, así como la gestión de sus emociones y dinámicas interpersonales en el trabajo.

Desde una perspectiva más específica de la Economía de la Empresa, los principales objetivos de los RRHH se basan en la toma de decisiones y acciones que influyen en la relación entre la empresa y sus empleados. Estas decisiones y acciones se toman con el fin de alcanzar los objetivos empresariales.

Estos objetivos se asocian desde el momento en que la actividad empresarial genera una relación laboral entre la empresa y una o más personas, independientemente de las características de la empresa (tamaño, situación mercantil, actividad que desarrolla, etc.) y de la relación establecida (tipo de contrato, duración del servicio, remuneración, etc.) (Gallastegi & Rodríguez, 2013).

Aunque esta función ha estado implícitamente presente en las tareas de gestión desde los primeros tiempos de la actividad empresarial, el desarrollo de las técnicas y conocimientos científicos dentro de la gestión de los RRHH no se produce hasta finales del siglo XIX como se ha mencionado anteriormente. Este hecho se produce como resultado de la evolución experimentada por la gestión de empresas a partir de la implantación de las fábricas como base del sistema productivo (Gallastegi & Rodríguez, 2013). La evolución de los objetivos de los RRHH se asocia a los cambios que se van produciendo en el sistema socioeconómico en el que está inserta.

Por otro lado, el desarrollo de los RRHH no se produce de manera uniforme en todo el tejido empresarial, sino que se circunscribe principalmente a las grandes empresas. En este tipo de empresas, su función requiere una mayor atención por parte de la dirección, ya que la complejidad inherente a su gestión y su grado de incidencia sobre la actividad empresarial es sustancialmente mayor que en las pequeñas y medianas empresas.

Además, la evolución de las distintas características de los RRHH en la empresa no ha sido idéntica en todos los países ni en el tiempo ni en los contenidos de la función, ya que los niveles de desarrollo industrial, social, político y económico han sido dispares entre los diferentes países, aunque es cierto que, en términos generales, las líneas de avance en los países más desarrollados han sido bastante similares.

Los RRHH también hacen alusión a la suma de conocimientos, experiencias, destrezas y dedicación que los trabajadores contribuyen a la empresa. Esta perspectiva incluye elementos menos tangibles como las emociones y la conexión emocional que los trabajadores establecen con su entorno laboral. En esencia, los RRHH son un pilar clave para la operatividad y prosperidad de cualquier organización, pues no solo atienden las demandas administrativas, sino que también son determinantes en el desarrollo del talento humano y en forjar un clima laboral constructivo (Martín Sierra, 2011)

Sus objetivos principales se centran en dirigir eficazmente el conjunto de habilidades y competencias de los trabajadores para alcanzar las metas establecidas. Es crucial para la empresa asegurar que cada miembro del equipo disfrute de un entorno

laboral enriquecedor y positivo. Esto implica no solo una adecuada selección y capacitación, sino también el desarrollo de un clima organizacional que promueva la satisfacción y el bienestar del personal, lo que a su vez se refleja en una mayor productividad y logro de los objetivos corporativos (Martín Sierra, 2011).

La Función de Recursos Humanos (FRRHH) se centra en tres aspectos clave: la eficiencia económica, la eficiencia social y la coherencia de sus acciones.

La eficiencia económica en el ámbito de los recursos humanos se relaciona con la mejora de los programas de rendimiento de una empresa (producción y distribución de bienes y servicios), a través de las aportaciones de los empleados, basándose en los principios de productividad laboral (óptima relación entre el rendimiento y el uso del factor) y economía de este (máximo rendimiento con costes mínimos). Además, la eficiencia económica abarca la consecución de metas de calidad (en las aportaciones de los individuos, en la organización y en la imagen externa) y la adquisición del nivel de flexibilidad y adaptabilidad necesario para contribuir de manera continua a la eficacia y eficiencia de la organización (Gallastegi & Rodríguez, 2013).

La eficiencia social se refiere a la mejora del nivel de satisfacción de las expectativas, necesidades e intereses de las personas de la organización y, en cierta medida, de la sociedad en general. Esto se reflejará en el nivel de compromiso e identificación de los empleados con su organización, en su satisfacción, motivación y contribución efectiva a la misma (productividad, calidad y adaptabilidad).

El objetivo de integración aborda tres aspectos diferentes:

1. La necesidad de que las políticas y objetivos de recursos humanos sean coherentes con el resto de las políticas de la empresa, es decir, que la capacidad laboral que se pretenda adquirir en cada momento y la forma en que se quiera obtener, se alinee con lo que la organización necesita para desarrollar sus estrategias y objetivos generales.
2. La necesidad de lograr la debida coherencia entre las diferentes políticas y objetivos de recursos humanos (selección, formación, retribución, etc.).
3. La búsqueda conjunta de la eficacia económica y social, y la integración de los objetivos de la organización y de sus trabajadores.

A modo de conclusión, el área de Recursos Humanos (RRHH) en una empresa juega un papel crucial en la supervisión y gestión de los empleados. Aunque originalmente se centraba en tareas administrativas, ha evolucionado para abarcar responsabilidades que se alinean con la visión y estrategia del equipo humano. La gestión de RRHH aborda todos los procesos relacionados con los trabajadores, desde su incorporación hasta el final de su relación laboral, y va más allá de las tareas burocráticas para incluir la promoción del compromiso y la motivación de los trabajadores. Los objetivos de RRHH, que ha estado presente desde los primeros tiempos de la actividad empresarial, ha evolucionado en paralelo a los cambios en el sistema socioeconómico. Aunque esta evolución no ha sido uniforme en todas las empresas, es especialmente relevante en las grandes empresas debido a la complejidad inherente a su gestión.

2.2 La digitalización de los RRHH en España

La administración efectiva del talento humano representa una meta clave para las corporaciones, y la evolución de las “tecnologías emergentes” está transformando el enfoque para alcanzar dicha meta. La era de la revolución tecnológica y digital ha inaugurado una era para el ámbito empresarial, marcada por el acceso simplificado a una vasta cantidad de datos e información. Además, como resultado de esta revolución, han surgido múltiples herramientas que ofrecen un valor significativo para la administración empresarial y, en especial, para la gestión de recursos humanos (Fernández Rodríguez et al., 2018). Uno de los desafíos más significativos para las empresas en la gestión de su capital humano es determinar cómo captar y procesar la extensa cantidad de datos e información de manera que resulte beneficiosa para la organización. Igualmente, cómo utilizar las tecnologías digitales modernas para lograr estos fines (Fernández Rodríguez et al., 2021).

Según el estudio realizado por la Organización Internacional de Directivos de Capital Humano (DHC) sobre la digitalización de los RRHH en España, han identificado cinco dimensiones tecnológicas que engloban la gestión de los RRHH para medir su grado de digitalización. Estas funciones se refieren a:

- **Big Data y Analytics:** Esta dimensión se enfoca en analizar múltiples cantidades de datos complejos para identificar patrones de comportamiento en los empleados.

- **Cloud Computing (Computación en la nube):** Se refiere a la tecnología que posibilita el acceso remoto a software a través de internet. También permite compartir y acceder a información de forma remota.
- **Redes Sociales:** En este sentido, las redes sociales son plataformas web que facilitan la interacción virtual entre sus miembros con el fin de compartir contenidos.
- **Movilidad:** Este concepto se relaciona con el uso generalizado de dispositivos móviles, especialmente smartphones.
- **Gamificación:** Este término, de origen anglosajón, implica la utilización de juegos digitales, como la realidad virtual o los juegos multimedia, con diversos propósitos digitales, como la realidad virtual o los juegos multimedia, para diversos fines.

A través de los siguientes capítulos, el trabajo explica de forma práctica cómo estos pilares tecnológicos complementan las políticas que engloban las funciones de los recursos humanos en la empresa. Se han definido nueve grupos de políticas ligados a las dimensiones tecnológicas para medir el grado de digitalización:

1. **Selección de talento:** Esta área abarca todas las actividades relacionadas con el reclutamiento, selección e incorporación de las nuevas personas contratadas.
2. **Formación y desarrollo:** Aquí se incluye el desarrollo personal y profesional, la creación de mapas de procesos, el fomento de carreras profesionales y la valoración del desempeño.
3. **Administración de personal:** Esta función se encarga de gestionar aspectos como nóminas, vacaciones y bajas.
4. **Compensación y beneficios:** Engloba todo lo relacionado con la retribución y los procesos de compensación.
5. **Salud laboral:** Se ocupa de la prevención de riesgos laborales.
6. **Human Resources Business Partner (HrbP):** Esta función implica comunicar las políticas y operativas de recursos humanos a cada área funcional dentro de la empresa.
7. **Digitalización:** Su objetivo es gestionar la transición digital de las organizaciones, identificando necesidades y creando las herramientas necesarias.

8. **Employer branding:** Se centra en potenciar la imagen de marca y el compromiso tanto con los empleados actuales como con sus potenciales candidatos.
9. **Relaciones laborales:** Aunque no se menciona en el texto original, esta área se encarga de las relaciones entre la empresa y los empleados, incluyendo negociaciones colectivas y resolución de conflictos.

La gestión de personas es uno de los objetivos primordiales de las empresas, y el avance de las “nuevas tecnologías” está transformando la forma en que se logra este objetivo. La revolución tecnológica y digital ha creado un nuevo escenario empresarial, caracterizado por la disponibilidad de un gran volumen de información y datos. Además, esta revolución ha dado lugar a la aparición de numerosas herramientas que pueden ser de gran utilidad tanto para la gestión general de la empresa como para la gestión de recursos humanos en (Fernández Rodríguez et al., 2018).

Un desafío importante al que se enfrentan las empresas en relación con la gestión de sus recursos humanos es cómo capturar y procesar eficazmente el enorme volumen de datos e información disponible. Asimismo, es fundamental comprender cómo aprovechar las nuevas tecnologías digitales para alcanzar estos objetivos.

En resumen, la gestión eficaz del talento humano es un objetivo esencial para las empresas, y el avance de las “tecnologías emergentes” está cambiando la forma en que se logra este objetivo. La revolución tecnológica y digital ha dado lugar a una nueva era en el mundo empresarial, caracterizada por el acceso simplificado a una gran cantidad de datos e información. Como resultado de esta revolución, han surgido numerosas herramientas que ofrecen un valor significativo para la gestión empresarial y, en particular, para la gestión de recursos humanos. Uno de los desafíos más importantes para las empresas en la gestión de su capital humano es determinar cómo capturar y procesar de manera efectiva la gran cantidad de datos e información disponible, y cómo utilizar las modernas tecnologías digitales para lograr estos fines. Las cinco dimensiones tecnológicas identificadas por la Organización Internacional de Directivos de Capital Humano (DHC) en su estudio sobre la digitalización de los RRHH en España, junto con las nueve políticas definidas, proporcionan un marco sólido para medir el grado de

digitalización y para entender cómo estas tecnologías pueden complementar las funciones de los recursos humanos en la empresa.

2.3 Modelos predictivos en RRHH y su impacto económico en las empresas

El “machine learning” (ML) es una rama de la inteligencia artificial que se aplica en el ámbito de los Recursos Humanos (RRHH) para analizar grandes volúmenes de datos y mejorar la toma de decisiones. En RRHH, el ML puede utilizarse para optimizar procesos como la selección de personal, el desarrollo del talento y la retención de empleados.

Una de las aplicaciones más innovadoras del ML en RRHH es la predicción de la renuncia voluntaria de los empleados. Mediante algoritmos de ML, las empresas pueden analizar patrones en los datos de los empleados, como su rendimiento, satisfacción laboral, trayectoria profesional y otros factores relevantes. Estos modelos pueden identificar señales de alerta que indican una mayor probabilidad de que un empleado decida dejar la empresa (Navarro,2022).

Por ejemplo, un sistema de ML puede examinar la frecuencia de ausencias de un empleado, su participación en programas de capacitación, la regularidad de sus promociones y cambios en su nivel de productividad. Al correlacionar estos datos con patrones de renuncia previos, el ML puede prever con cierta precisión quién podría estar considerando la renuncia.

Esta capacidad predictiva es valiosa para los departamentos de RRHH porque permite intervenir de manera proactiva y reducir el impacto económico de la rotación. Los modelos predictivos pueden identificar a los empleados que tienen más probabilidades de abandonar la organización. Esto permite a los gerentes y al departamento de RR.HH. intervenir tempranamente, ya sea a través de conversaciones de retención, ajustes salariales, oportunidades de desarrollo profesional, o mejoras en el equilibrio entre la vida laboral y personal. como analizaremos durante los siguientes capítulos (Manatal,2023).

Por otro lado, al identificar de manera temprana a los candidatos que probablemente permanecerán en la organización a largo plazo, las empresas pueden

optimizar la eficiencia de sus procesos de contratación y reducir los costos asociados con la rotación de personal. La rotación de personal puede afectar negativamente la productividad, ya que los empleados recién contratados necesitan un período de adaptación y aprendizaje en sus nuevos roles. Al reducir la rotación de personal, las empresas pueden garantizar la continuidad de altos niveles de productividad.

Existen teorías de retención de empleados que pueden ser aplicadas a los modelos predictivos en Recursos Humanos de varias maneras. Por ejemplo, los modelos predictivos en Recursos Humanos pueden incorporar la teoría de las necesidades de Maslow al analizar datos que reflejen cuán satisfechas están las necesidades de los empleados. Si los datos indican que las necesidades básicas de los empleados, como la seguridad laboral o las condiciones de trabajo, no se están satisfaciendo, el modelo podría prever un aumento en la rotación de personal.

Otra de las teorías que puede ser aplicada a un modelo predictivo es la teoría del compromiso organizacional de Meyer y Allen al examinar datos que reflejen el grado de compromiso de los empleados con la organización. Si los datos sugieren que los empleados tienen un alto nivel de compromiso afectivo, lo que significa que tienen un fuerte lazo emocional con la organización, el modelo podría anticipar una disminución en la rotación de personal (Ingenia, 2020).

En conclusión, el “machine learning” (ML) es una herramienta poderosa en el campo de los Recursos Humanos (RRHH), permitiendo analizar grandes cantidades de datos y mejorar la toma de decisiones. Su aplicación en procesos como la selección de personal, el desarrollo del talento y la retención de empleados ha demostrado ser altamente efectiva. Una de las innovaciones más destacadas es la capacidad de predecir la renuncia voluntaria de los empleados, lo que permite a las empresas intervenir de manera proactiva y minimizar el impacto económico de la rotación de personal. Además, la identificación temprana de los candidatos que probablemente permanecerán en la organización a largo plazo optimiza la eficiencia de los procesos de contratación y reduce los costos asociados con la rotación de personal. Finalmente, la incorporación de teorías de retención de empleados en los modelos predictivos, como la teoría de las necesidades de Maslow y la teoría del compromiso organizacional de Meyer y Allen, permite un

análisis más profundo y preciso de los datos de los empleados, mejorando aún más la eficacia de estos modelos

3 INVESTIGACIÓN MEDIANTE ANÁLISIS DE DATOS

3.1 Preparación del conjunto de datos

Se obtuvo un “dataset” cuyos datos hacían referencia a distintas personas que están presentes en el mundo laboral, indicando sobre las mismas, variables que podrían influir en la decisión de un empleado de dejar su trabajo. Se trata de un conjunto de datos ficticio que ha sido creado por “data scientists” de “International Business Machines (IBM)”, a fin de exponer los principales factores que conducen a la super rotación de empleados. La variable “target”, es decir, la que se quiere predecir va a ser “Attrition”, que indicará si el empleado acaba finalmente abandonando la empresa o no. Toma valores de “sí” o “no” (HR-Employee-Attrition,2017).

El conjunto de datos utilizado se compone de 1471 registros de personas y 35 variables. Sin embargo, después de importar el “dataset”, se han eliminado aquellas variables que se consideraba que no aportan información o que podrían ser un obstáculo de cara a la predicción del objetivo.

Entre ellas se identifican variables cuyos valores eran únicos para cada persona: “EmployeeCount” o “Employee number”, que eran números identificativos del empleado. Variables que, por el contrario, tomaban el mismo valor para todas las observaciones “Over18” o “StandardHours”. O variables cuyo significado de sus valores no quedaba claro en la descripción del dataset y por tanto podrían generar confusión: “Daily rate”, “Hourly rate”, “Distance from home”, “Monthly rate”, “Percent salary hike” o “StockOptionLevel”.

A continuación, se ilustra un cuadro resumen, donde aparecen todas las variables que finalmente se utilizaran para realizar el modelo, con una descripción, el tipo de variables que son, el rango de valores que toman y ciertos comentarios que se han considerado relevantes.

Tabla 1: Resumen de variables

Variable	Descripción	Tipo	Rango de valores	Comentarios
Age	Edad del trabajador	Numérica	18-60	La edad del trabajador influye a la hora de tomar la decisión de dejar una empresa
Attrition	Decisión de abandonar o no la empresa	Dicotómica	Sí: Decide abandonar la empresa No: Decide no abandonar la empresa	Variable que queremos predecir
BusinessTravel	Nivel de viajes por trabajo	Factor	Non_Travel, Travel_Rarely o Travel_Frequently	Según la persona el viajar por trabajo puede ser un beneficio o un punto negativo
Department	Departamento en el que trabaja	Factor	Sales, Research & Development o Human Resources	Hemos agrupado el departamento de trabajo en estas 3 categorías para simplificar. Research & Development incluye todos los campos que no son "Sales" o "Human Resources"
Education	Nivel de estudios	Catagórica ordinal	1: "Below College" 2: "College"	Una persona con mayor nivel de estudios tiene

			3: "Bachelor" 4: "Master" 5: "Doctor"	más posibilidades de encontrar un nuevo trabajo
EducationField	Disciplina de estudios	Factor	"Human Resources", "Life Sciences", "Marketing", "Medical", "Other", "Technical Degree"	Hay disciplinas de estudios donde hay más oportunidades de trabajo que en otras
EnvironmentSatisfaction	Satisfacción con el ambiente de trabajo	Categórica ordinal	1: "Low" 2: "Medium" 3: "High" 4: "Very high"	Toma valores de 1 al cuatro, siendo 1 poco satisfecho y 4 muy satisfecho
Gender	Género	Dicotómica	"Female" o "Male"	Hace referencia a si es trabajador es de sexo femenino o masculino
JobInvolvement	Grado de compromiso laboral	Categórica ordinal	1: "Low" 2: "Medium" 3: "High" 4: "Very high"	Toma valores del 1 al 4 de menor a mayor según el grado de compromiso del empleado con su puesto de trabajo
JobLevel	Nivel de puesto de trabajo	Categórica ordinal	1: "Low" 2: "Medium" 3: "High" 4: "Very high"	Se refiere al nivel jerárquico o de responsabilidad del trabajador en la empresa.

				Siendo 1 menor grado de responsabilidad y 4 el máximo.
JobRole	Puesto de trabajo	Factor	"Healthcare Representative", "Human Resources", "Laboratory Technician", "Manager", "Manufacturing Director", "Research Director", "Research Scientist", "Sales Executive" o "Sales Representative"	Hace referencia a diferentes puestos de trabajo que el empleado puede tener en dentro de la organización
JobSatisfaction	Satisfacción laboral	Categórica ordinal	1: "Low" 2: "Medium" 3: "High" 4: "Very high"	De menor a mayor indicando el grado de satisfacción con la empresa en general
MaritalStatus	Estado civil	Factor	"Divorced", "Married" y "Single"	Indica si el trabajador está divorciado, casado o soltero.
MonthlyIncome	Salario mensual	Numérica	1000-20000	Indica el salario mensual del empleado

NumCompaniesWorked	Número de compañías en las que ha trabajado el empleado	Numerica	0-9	Hace referencia al número de compañías en las que ha trabajado el empleado
OverTime	Realización de horas extra o no	Dicotómica	“Sí” o “No”	Toma valor “Sí” si el empleado realiza horas extra de manera recurrente o “No” en caso contrario.
PerformanceRating	Evaluación del desempeño	Categórica ordinal	1: “Low” 2: “Good” 3: “Excellent” 4: “Outstanding”	Se refiere a la evaluación sistemática que se realiza a un empleado para otorgarles una calificación a su rendimiento. Toma valores de 1 a 4 de menos a mayor.
RelationshipSatisfaction	Nivel de satisfacción con compañeros de trabajo	Categórica ordinal	1: “Low” 2: “Medium” 3: “High” 4: “Very high	Toma valores del 1 al 4. Siendo la 1 mala relación con los compañeros de trabajo y 4 muy buena relación
TotalWorkingYears	Número total de años trabajados	Numérica	0-40	Los años que lleve el empleado en la empresa pueden influir

				al tomar la decisión de abandonarla
TrainingTimeLastYear	Número de veces en las que un trabajador ha realizado un curso de formación en el último año	Numérica	0-6	La inversión en formación de los trabajadores influye en su decisión de dejar la empresa
WorkLifeBalance	Nivel de conciliación del trabajo con la vida personal	Categórica ordinal	1: "Low" 2: "Medium" 3: "High" 4: "Very high"	Establece el nivel de equilibrio entre vida profesional y laboral
YearsAtCompany	Número de años en la compañía	Numérica	0-40	El número de años que permanecen los trabajadores en la misma empresa tiene relación con el nivel de rotación de la misma
YearsInCurrentRole	Número de años en el puesto actual	Numérica	0-18	Pensamos que existe una relación entre el número de años en el mismo puesto y la satisfacción en el trabajo
YearsSinceLastPromotion	Número de años el último ascenso del empleado	Numérica	0-15	Creemos que la frecuencia de ascenso en una empresa está relacionada

				con la satisfacción de los trabajadores
YearsWithCurrentManager	Número de años con el mismo jefe	Numérica	0-17	Hace referencia a los años que un empleado lleva con un mismo jefe. Pensamos que puede verse tanto como un aspecto positivo como negativo por parte del empleado.

Fuente 1: Elaboración propia

3.2 Análisis descriptivo

Para comenzar se realiza un análisis descriptivo del dataset. Se ha utilizado la herramienta “R-studio” durante todo el trabajo. Se trata de un entorno integrado de desarrollo (IDE) para el lenguaje de programación R. Es una herramienta muy útil para el análisis, modelado estadístico y visualización de datos.

Primeramente, se llama a la función “summary” para obtener un resumen con todos los estadísticos de cada variable. Más adelante, se realizará el tratamiento a ciertas variables que lo necesitan. Para estudiar la correlación entre variables categóricas y no únicamente las numéricas, utilizaremos la técnica conocida como "one-hot encoding". Este método permite convertir las variables categóricas en un formato que puede ser interpretado numéricamente. Consiste en crear nuevas columnas a partir de la variable categórica original, generando una columna nueva para cada categoría única dentro de dicha variable (One hot encoding, 2023). Estas nuevas columnas son binarias, es decir, solo contienen los valores 0 o 1. En cada fila de datos, se asigna un valor de 1 a la columna correspondiente a la categoría presente, mientras que las demás columnas reciben un valor de 0. Sin embargo, para evitar problemas de multicolinealidad, que ocurren cuando hay

una alta correlación entre variables predictoras, es necesario eliminar una de estas columnas. La columna eliminada se designa como la variable base. No hay un procedimiento específico para seleccionar esta variable base; puede ser cualquiera de las categorías originales. Esta estrategia ayuda a evitar redundancias en el modelo y facilita una interpretación más clara y significativa de los coeficientes en técnicas de regresión. Por ejemplo, si tenemos una variable llamada "color" con valores "rojo", "azul" y "verde", y seleccionamos "verde" como la base, crearemos dos columnas binarias: "columna_azul" y "columna_rojo". Si un registro tiene el valor "azul", la columna "columna_azul" tendrá un 1, mientras que "columna_rojo" tendrá un 0. Si el valor es "rojo", "columna_rojo" tendrá un 1 y "columna_azul" un 0. Si el valor es "verde", ambas columnas tendrán un valor de 0 (Tovar, 2022). Este método permite que las variables categóricas sean utilizadas eficazmente en modelos de análisis de datos y aprendizaje automático.

Como se acaba de mencionar, para implementar “One-hot encoding” hay que seleccionar una categoría base en cada variable, la categoría base se selecciona de manera aleatoria, lo indicaremos a continuación:

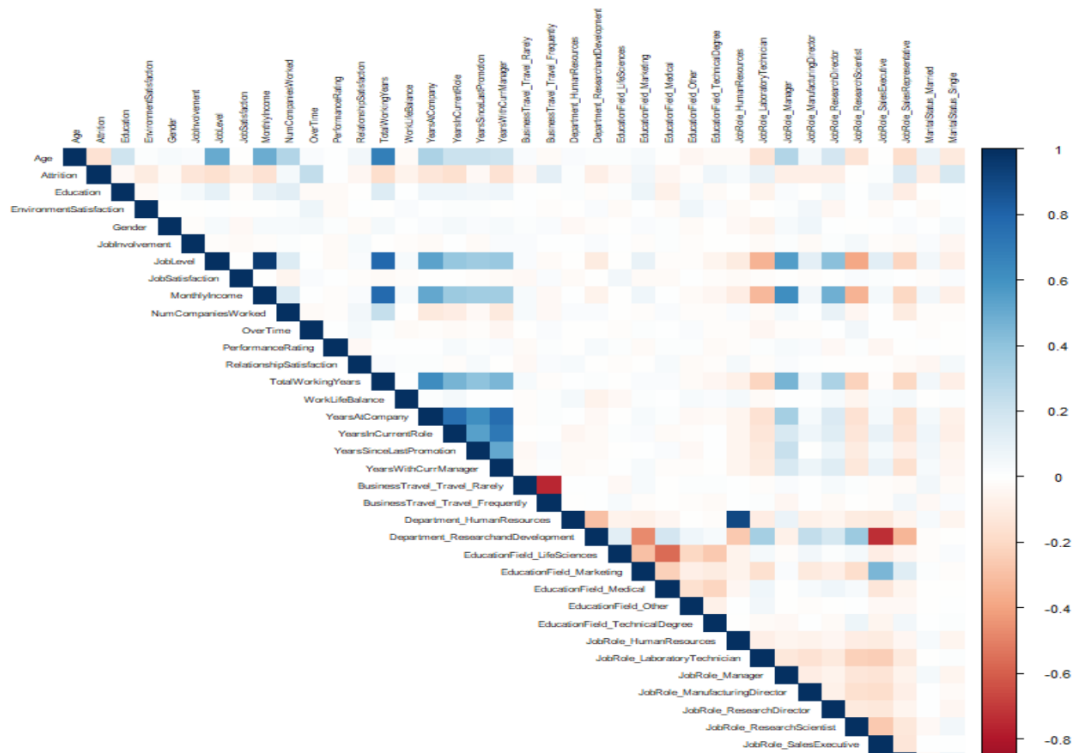
- Para la variable “BusinessTravel” hemos utilizado como base ‘Non_Travel’
- Para la variable “Department” hemos utilizado como base ‘Sales’
- Para la variable “EducationField” hemos utilizado como base ‘Human Resources’
- Para la variable “JobRole” hemos utilizado como base ‘Healthcare Representative’
- Para la variable “MaritalStatus” hemos utilizado como base ‘Divorced’

A su vez, hemos modificado las variables dicotómicas para que tomen valores 0 y 1 en vez de “Yes” y “No”

- Para la variable “Attrition” aparecerá un 1 en los valores “Yes” y un 0 en los valores “No”.
- Para la variable “Gender” aparecerá un 1 en los valores “Female” y un 0 en los valores “Male”.
- Para la variable “OverTime” aparecerá un 1 en los valores “Yes” y un 0 en los valores “No”.

A continuación, se eliminarán los posibles valores nulos del “dataset” para no arrastrarlos durante los siguientes pasos, dado que pueden generar errores. Una vez tenemos el “dataset” sin datos nulos y con las nuevas variables categóricas tratadas, nos disponemos a calcular la matriz de correlaciones. A través de la matriz se observan las combinaciones de correlaciones entre todas las variables presentes en nuestro conjunto de datos. A la hora de interpretarla se presta atención a la nitidez tanto de color azul como del rojo en la matriz de correlaciones. Cuanto más nítido sea el color azul del recuadro en cuestión mayor será la correlación positiva de ambas variables y cuanto más fuerte sea el color rojo del recuadro mayor será la correlación negativa.

Figura 1: Mapa de calor, correlaciones



Fuente 2: Elaboración propia

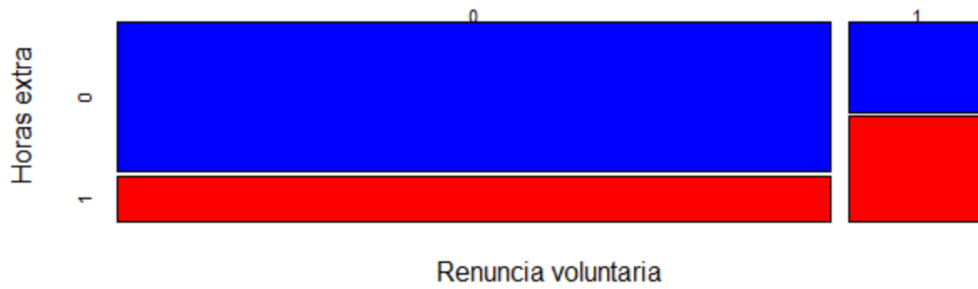
No obstante, el gran número de variables que presenta el “dataset” hace que resulte complicado observar de manera precisa los valores exactos de las correlaciones en el gráfico. Asimismo, se decide obtener las variables que tienen la mayor correlación positiva y la mayor correlación negativa con nuestra variable objetivo.

Como resultado se obtiene que la variable que representa el abandonar la empresa está directamente muy relacionada con factores como “OverTime” , “MaritalStatus_Single” y “BusinessTravel_Travel_Frequently”, lo cual resulta coherente, puesto que , el realizar horas extra en el trabajo, el tener un trabajo que demande al trabajador viajar mucho y el estar soltero son variables que influyen directamente a la hora de que un empleado decida abandonar la empresa en la que trabaja de manera voluntaria. Según la matriz obtenida cuantas más horas extra realice el trabajador, más se le exija viajar o si esta soltero más aumenta la probabilidad de que renuncie de manera voluntaria. A su vez la variable “Attrition” esta inversamente relacionada con variables como “TotalWorkingYears”, “Job level”, “MonthlyIncome”. Es decir, que cuantos más años lleve el empleado en el mundo laboral, mejor posición tenga en la empresa, o mayor sea el salario mensual, menor es la probabilidad que abandone la empresa voluntariamente.

Aunque estas sean las variables que presentan una mayor correlación, ninguno de los valores sobrepasa el 0,5. Sin embargo, queríamos estudiar más a fondo el comportamiento de la variable objetivo con las variables “OverTime” y “TotalWorkingYears”, que son aquellas con las que tiene una correlación positiva y negativas más fuerte respectivamente. Se realizará un gráfico de mosaico para comparar las variables “renuncia voluntaria” y “Horas extra”, dado que ambas son variables categóricas.

A través de este gráfico de mosaico se observa claramente como en el segmento en el que la variable objetivo “Renuncia voluntaria” toma valor 1 hay una mayor presencia de la variable “Horas extra” tomando valor 1, que cuando la variable “Renuncia voluntaria” toma valor 0. Lo cual tiene sentido, puesto que el aumento de horas extra en un puesto de trabajo son uno de los motivos impulsores del abandono voluntario de los empleados de las empresas

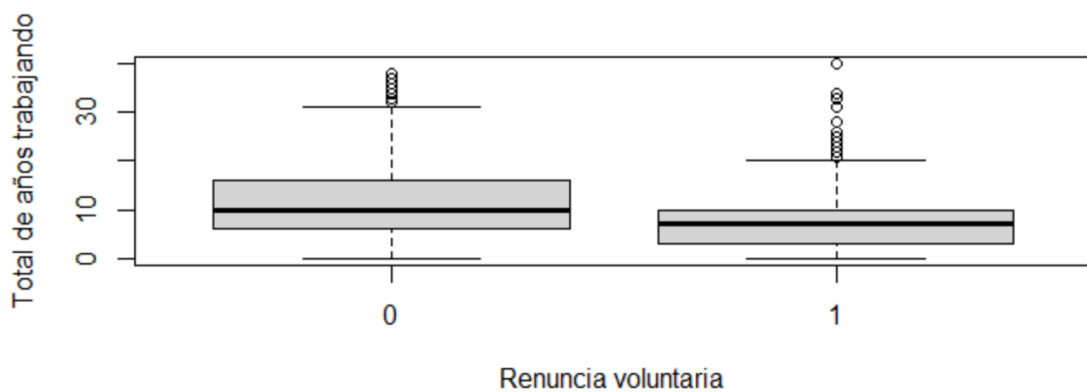
Figura 2: Renuncia voluntaria Vs Horas extra



Fuente 3:Elaboración propia

A continuación, se realiza un diagrama de cajas para observar la comparativa entre las variables “Renuncia voluntaria” y “Número total de años en el mundo laboral”. Se aprecia como en el momento en el que la variable objetivo “Renuncia voluntaria” toma valor 1 la media del total de años trabajados es menor que cuando toma valor 0. Esto pone de manifiesto que los trabajadores que abandonan de manera voluntaria su puesto de trabajo llevan un menor número de años en el mundo laboral. Lo cual tiene sentido, dado que durante los primeros años en el mundo laboral los empleados tienden más a rotar por distintas empresas hasta que encuentran aquella que más encaja con sus aptitudes y preferencias.

Figura 3: Renuncia voluntaria Vs Número total de años en el mundo laboral



Fuente 4: Elaboración propia

En resumen, se realiza un análisis descriptivo de un conjunto de datos usando R-studio. Se emplea la función "summary" para obtener estadísticas iniciales y se usa "one-hot encoding" para transformar variables categóricas en binarias, eliminando una categoría base para evitar multicolinealidad. Se ajustan variables dicotómicas a valores 0 y 1 y se eliminan valores nulos.

Se calcula una matriz de correlaciones para identificar las relaciones entre variables. La renuncia voluntaria de empleados se correlaciona positivamente con horas extra, viajar frecuentemente y estar soltero, y negativamente con los años trabajados, nivel de empleo y salario mensual.

3.3 Modelo explicativo

Seguidamente, se realiza un modelo explicativo. Como ya se ha mencionado anteriormente, la variable objetivo es el campo dicotómico que hace referencia a si el empleado renuncia de manera voluntaria en la empresa en la que está contratado o si no. Por lo tanto, se utiliza un algoritmo de regresión logística (LOGIT) que permita llevar a cabo una clasificación binaria.

Esto permitirá estimar cual es la probabilidad de que un trabajador renuncie a su puesto de trabajo o no.

A través de este análisis, se identificarán las variables que más influyen en la toma de decisión de un empleado de abandonar una empresa.

Figura 4: Salida modelo explicativo

```
Call:
glm(formula = Attrition ~ ., family = binomial, data = datos)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7523  -0.5066  -0.2689  -0.1060   3.5847

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.000e+00  1.809e+00  1.105  0.26897
Age          -3.328e-02  1.336e-02  -2.491  0.01275 *
Education    1.639e-02  8.636e-02  0.190  0.84948
EnvironmentSatisfaction -4.185e-01  8.098e-02  -5.168  2.37e-07 ***
Gender       -3.876e-01  1.810e-01  -2.142  0.03221 *
JobInvolvement -5.033e-01  1.206e-01  -4.174  2.99e-05 ***
JobLevel     1.716e-02  3.107e-01  0.055  0.95596
JobSatisfaction -4.076e-01  7.951e-02  -5.126  2.95e-07 ***
MonthlyIncome -8.784e-06  8.022e-05  -0.109  0.91281
NumCompaniesworked 1.814e-01  3.798e-02  4.778  1.77e-06 ***
OverTime     1.949e+00  1.893e-01  10.294  < 2e-16 ***
PerformanceRating -6.005e-02  2.440e-01  -0.246  0.80559
RelationshipsSatisfaction -2.424e-01  8.078e-02  -3.001  0.00269 **
TotalWorkingyears -5.524e-02  2.907e-02  -1.900  0.05745 .
workLifeBalance -3.582e-01  1.211e-01  -2.957  0.00311 **
YearsatCompany  8.793e-02  3.850e-02  2.284  0.02237 *
YearsInCurrentRole -1.449e-01  4.449e-02  -3.256  0.00113 **
YearsSinceLastPromotion 1.767e-01  4.148e-02  4.260  2.04e-05 ***
YearsWithCurrManager -1.322e-01  4.634e-02  -2.854  0.00432 **
BusinessTravel_Travel_Rarely 8.390e-01  3.624e-01  2.315  0.02060 *
BusinessTravel_Travel_Frequently 1.666e+00  3.905e-01  4.268  1.97e-05 ***
Department_HumanResources -1.245e+01  3.800e+02  -0.033  0.97387
Department_ResearchandDevelopment 4.059e-01  1.043e+00  0.389  0.69705
EducationField_Lifesciences -7.611e-01  7.800e-01  -0.976  0.32918
EducationField_Marketing -3.008e-01  8.275e-01  -0.363  0.71625
EducationField_Medical -8.867e-01  7.808e-01  -1.136  0.25608
EducationField_Other -8.072e-01  8.400e-01  -0.961  0.33659
EducationField_TechnicalDegree 6.837e-02  7.998e-01  0.085  0.93188
JobRole_HumanResources 1.418e+01  3.800e+02  0.037  0.97023
JobRole_LaboratoryTechnician 1.467e+00  4.764e-01  3.079  0.00207 **
JobRole_Manager 3.668e-01  8.768e-01  0.418  0.67570
JobRole_ManufacturingDirector 2.831e-01  5.241e-01  0.540  0.58907
JobRole_ResearchDirector -9.022e-01  9.617e-01  -0.938  0.34821
JobRole_ResearchScientist 5.719e-01  4.869e-01  1.175  0.24016
JobRole_SalesExecutive 1.410e+00  1.120e+00  1.259  0.20808
JobRole_SalesRepresentative 2.237e+00  1.177e+00  1.901  0.05732 .
MaritalStatus_Married 4.005e-01  2.522e-01  1.588  0.11227
MaritalStatus_Single 1.369e+00  2.570e-01  5.327  9.97e-08 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1298.58  on 1469  degrees of freedom
Residual deviance:  888.12  on 1432  degrees of freedom
AIC: 964.12

Number of Fisher Scoring iterations: 14
```

Fuente 5: Elaboración propia

En este primer modelo se utilizan todas las variables, con el fin de identificar las más influyentes para posteriormente ir acotando el modelo únicamente usando estas características de más peso.

Si se observan los p-valores de cada variable se es capaz de identificar aquellas variables a las que es más sensible un empleado a la hora de abandonar su trabajo. Cuanto más pequeño sea el p-valor más influyente son las variables. Puesto que, la probabilidad de estar cometiendo un error al rechazar la hipótesis nula, siendo dicha hipótesis que los resultados se deban al azar, es prácticamente inexistente. Es decir, que resulta más probable que las variables tengan un efecto significativo en la variable objetivo.

Entre ellas destacan:

- EnvironmentSatisfaction: Resulta evidente la influencia que tiene la satisfacción con el ambiente de trabajo sobre la toma de decisión a la hora de abandonar una organización. Si un empleado no está cómodo con su entorno de trabajo aumenta la probabilidad de renuncia voluntaria.
- JobInvolvement: El nivel de compromiso tanto a nivel profesional como personal que muestra un trabajador puede revelar intenciones de permanecer en la empresa a largo plazo o, todo lo contrario.
- JobSatisfaction: El grado de satisfacción con la empresa en general influye en la opinión de un trabajador sobre las misma. Cuanto mayor sea el grado de satisfacción general menor será la probabilidad de renuncia voluntaria
- NumCompaniesWorked: El número de compañías por las que ha pasado el empleado puede influir a la hora de renunciar a su puesto de trabajo actual. Por un lado, puede que un trabajador que haya pasado por muchas empresas haya podido adquirir una mayor experiencia y haya desarrollado más competencias. Como consecuencia, puede ser más demandado por otras empresas y por lo tanto más propenso a renunciar voluntariamente. No obstante, aquellas personas que hayan pasado por menos empresas pueden indicar un mayor grado de compromiso y búsqueda de estabilidad. Siendo menos propensos a abandonar un puesto de trabajo.
- OverTime: Esta variable hace referencia a las horas extra realizadas por el empleado que están ampliamente ligadas al aumento de probabilidad de renuncia voluntaria. El famoso “burnout”, es un término que hace referencia al agotamiento tanto mental como físico generado por el estrés laboral crónico
- RelationshipSatisfaction: La siguiente variable hace referencia a la satisfacción del empleado con sus compañeros o equipos de trabajo. Está demostrada la importancia de fomentar la creación de relaciones personales positivas en el entorno de trabajo para mejorar el rendimiento y satisfacción del empleado. Cuanto más favorables sean estas relaciones personales más disminuye la probabilidad de renuncia voluntaria

- **WorkLifeBalance:** La conciliación entre la vida laboral y la vida personal de empleado es un factor crucial a la hora de minimizar el grado de rotación en una empresa.
- **YearsInCurrentRole:** Esta variable hace referencia al número de años que el empleado lleva en su puesto actual. Según el modelo cuantos más años lleva un trabajador en el puesto, disminuye la probabilidad de que renuncie. Esto puede deberse a muchos factores como la comodidad, la lealtad, o el no querer rechazar un buen puesto, entre otros.
- **YearsSinceLastPromotion:** El número de años que han pasado desde la última vez que el empleado tuvo un ascenso. Se aprecia como a medida que los años sin ser promocionado aumentan, el trabajador es más propenso a abandonar la empresa. Resulta coherente, dado que el no tener metas a largo plazo o verse estancado pueden disminuir la satisfacción del empleado con su puesto de trabajo.
- **YearsWithCurrManager:** Esta variable indica el número de años que el trabajador lleva bajo el mando del mismo manager. El modelo sugiere que cuantos más años lleva el empleado sin cambiar de jefe más disminuyen las probabilidades de renuncia voluntaria. Puede deberse a factores ya mencionados anteriormente como son lealtad, comodidad o buena relación con su superior entre otros.
- **BusinessTravel_Travel_Frequently:** El modelo pone de manifiesto que la variable que indica la frecuencia con la que el trabajador viaja está directamente relacionada con el abandono voluntario del puesto de trabajo. Es decir, que el tener un trabajo que obligue al empleado a viajar mucho suele considerarse un factor negativo.
- **MaritalStatus_Single:** Esta variable hace referencia a ser un trabajador cuyo estado civil sea soltero. Se observa como aquellos empleados solteros tienen una mayor tendencia a renunciar de manera voluntaria a un empleo. Esta relación puede generarse, dado que, el empleado no tiene responsabilidades extra como puede ser la familia, generando una menor aversión al riesgo.

Una vez identificadas las variables más relevantes pasaremos a generar nuestro modelo predictivo.

3.4 Modelos predictivo

Para llevar a cabo las predicciones utilizamos un algoritmo predictivo de regresión logística llamado logit. El objetivo principal es ser capaz de predecir si un empleado va a abandonar su puesto de trabajo de manera voluntaria para tomar las medidas necesarias a fin de evitar ese suceso y reducir el alto nivel de rotación en ciertas empresas.

DESCRIPCIÓN DEL ALGORITMO REGRESION LOGÍSTICA

El algoritmo regresión logística es una técnica de aprendizaje supervisado. Se utiliza cuando deseamos conocer la relación entre una variable dependiente dicotómica, es decir, que solo puede tomar dos valores, en este caso “renuncia voluntaria” y “renuncia no voluntaria”; y una serie de variables independientes (IBM,2021).

A través de este algoritmo se obtiene la probabilidad de que una observación pertenezca a una categoría u otra de nuestra variable “target” (García Pérez et al. 2010).

Figura 5: Algoritmo de regresión logística

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + b_2x_2 + \dots$$

Fuente 6: Elaboración propia

Primeramente, se realiza la partición del conjunto de datos en dos sets. Uno de ellos se utilizará como datos de entrenamiento para conseguir generar el modelo. Y el otro será el conjunto de validación que nos permitirá testar la calidad de nuestro modelo. En esta ocasión destinamos el 70% de los datos al conjunto de entrenamiento y el 30% restante al conjunto de validación, con 1029 y 441 entradas de registros de datos de empleados respectivamente.

MODELO 1

Para construir el primer modelo utilizaremos la función “glm”. En este primer modelo se utilizarán todas las variables independientes del “dataset”. En la imagen inferior se puede apreciar la salida de la función.

Figura 6: Salida función "glm" Modelo 1

```

Call:
glm(formula = Attrition ~ ., family = binomial(logit), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6925  -0.4757  -0.2678  -0.1075   3.6207

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.921e+00  2.093e+00   0.918  0.358662
Age          -2.748e-02  1.624e-02  -1.692  0.090639 .
Education    -1.472e-02  1.024e-01  -0.144  0.885640
Environment  -4.448e-01  9.755e-02  -4.559  5.13e-06 ***
Gender       -2.597e-01  2.181e-01  -1.191  0.233679
JobInvolvement -4.689e-01  1.466e-01  -3.200  0.001376 **
JobLevel     2.404e-02  3.879e-01   0.062  0.950581
JobSatisfaction -4.485e-01  9.831e-02  -4.562  5.08e-06 ***
MonthlyIncome -2.639e-05  1.000e-04  -0.264  0.791867
NumCompaniesWorked 1.912e-01  4.631e-02  4.128  3.66e-05 ***
OverTime     1.926e+00  2.316e-01  8.315  < 2e-16 ***
PerformanceRating -7.455e-02  3.015e-01  -0.247  0.804704
RelationshipSatisfaction -2.491e-01  9.905e-02  -2.515  0.011899 *
TotalWorkingYears -1.046e-01  3.771e-02  -2.775  0.005521 **
workLifeBalance -2.043e-01  1.491e-01  -1.370  0.170574
YearsAtCompany  7.708e-02  4.823e-02  1.598  0.109960
YearsInCurrentRole -1.109e-01  5.318e-02  -2.085  0.037042 *
YearsSinceLastPromotion 1.765e-01  4.963e-02  3.556  0.000376 ***
YearswithCurrManager -9.089e-02  5.624e-02  -1.616  0.106069
BusinessTravel_Travel_Rarely 6.660e-01  4.286e-01  1.554  0.120173
BusinessTravel_Travel_Frequently 1.672e+00  4.603e-01  3.632  0.000282 ***
Department_HumanResources -1.238e+01  4.766e+02  -0.026  0.979278
Department_ResearchandDevelopment 1.209e-01  1.101e+00  0.110  0.912516
EducationField_LifeSciences -5.245e-01  8.900e-01  -0.589  0.555612
EducationField_Marketing -2.970e-01  9.663e-01  -0.307  0.758570
EducationField_Medical -1.003e+00  8.956e-01  -1.120  0.262680
EducationField_Other -6.539e-01  9.735e-01  -0.672  0.501773
EducationField_TechnicalDegree 7.711e-02  9.148e-01  0.084  0.932827
JobRole_HumanResources 1.435e+01  4.766e+02  0.030  0.975986
JobRole_LaboratoryTechnician 1.631e+00  6.127e-01  2.662  0.007763 **
JobRole_Manager 1.328e+00  9.848e-01  1.348  0.177523
JobRole_ManufacturingDirector 6.904e-01  6.479e-01  1.065  0.286676
JobRole_ResearchDirector 1.727e-01  1.047e+00  0.165  0.868933
JobRole_ResearchScientist 5.406e-01  6.279e-01  0.861  0.389260
JobRole_SalesExecutive 1.353e+00  1.229e+00  1.101  0.270872
JobRole_SalesRepresentative 2.137e+00  1.309e+00  1.632  0.102650
MaritalStatus_Married 4.620e-01  3.127e-01  1.478  0.139504
MaritalStatus_Single 1.544e+00  3.209e-01  4.811  1.50e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 906.03  on 1028  degrees of freedom
Residual deviance: 606.45  on 991  degrees of freedom
AIC: 682.45

```

Fuente 7: Elaboración propia

Una vez entrenado el modelo 1, se utiliza la función “predict” con el fin de calcular las probabilidades predichas de renuncia voluntaria con los datos de trabajadores que se encuentran en el conjunto de datos de validación. Para poder evaluar así el rendimiento y funcionamiento de nuestro modelo.

A continuación, se calcula la matriz de confusión. Como siguiente paso, se establece el cut-off, que es el punto de corte a partir del cual una probabilidad de un empleado concreto será clasificada como “renuncia voluntaria” o “no renuncia voluntaria”.

Se ha seleccionado un umbral de corte de 0,3 para nuestro modelo. Este valor, al ser menor que 0,5, hace que el modelo sea más sensible a la detección de falsos negativos. La razón detrás de esta elección es que el costo de predecir falsos negativos es alto en nuestro caso de uso. Es decir, es más problemático predecir que un trabajador no se va a ir cuando en realidad sí lo hará.

Para entenderlo mejor, se piensa en un ejemplo concreto. Se imagina que se está utilizando un modelo para predecir si los empleados van a renunciar. Si el modelo predice que un empleado no va a renunciar (negativo) pero en realidad sí lo hace (falso negativo), no se tomarán las medidas necesarias para retener a ese empleado. Esto puede llevar a una pérdida de talento valioso y aumentar los costos de contratación y entrenamiento de nuevos empleados.

El objetivo principal es identificar posibles renunciaciones antes de que ocurran para poder intervenir a tiempo. Al usar un umbral más bajo como 0,3, el modelo considera más casos como positivos (posibles renunciaciones). Esto significa que el modelo será más propenso a predecir que un empleado podría renunciar, incluso si la certeza no es muy alta. Así, se reduce la cantidad de falsos negativos y aumentamos nuestras oportunidades de intervenir y evitar que los empleados se vayan.

A través de la matriz de confusión se comparan las predicciones del modelo con los valores reales de rotación de empleados. Se observa que el número de verdaderos positivos es 42, es decir, el modelo clasificó a 42 empleados como que pretendían abandonar su puesto de trabajo de manera voluntaria cuando realmente así era. Además, el valor de verdaderos negativos es de 328, son observaciones que el modelo clasificó de manera correcta como empleados que no pretendían renunciar de manera voluntaria cuando efectivamente no renunciaron. Asimismo, el valor de falsos positivos es 41, estos son empleados que el modelo clasificó incorrectamente como “renuncia voluntaria” cuando en realidad no renunciaron. Por último, el error que más perjudica, los falsos negativos, toma un valor de 30. Estos serían los empleados que el modelo clasificó incorrectamente como “No renuncia voluntaria” y en realidad si abandonaron su puesto de trabajo.

Figura 7: Matriz confusión Modelo 1

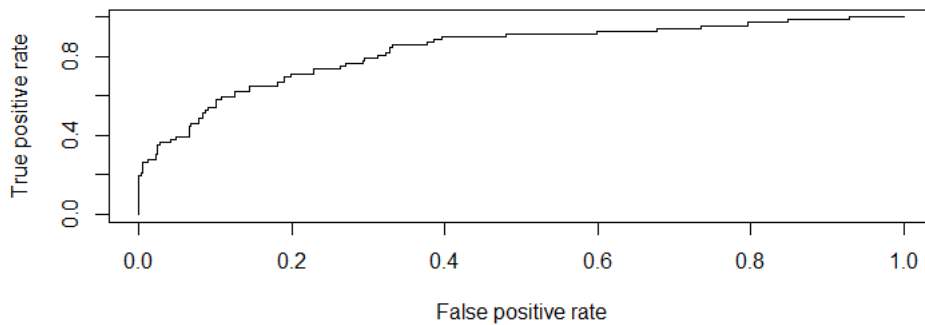
```
> matrizConfusion
      REAL
PREDICTED  0  1
FALSE    328 30
TRUE     41  42
```

Fuente 8: Elaboración propia

A continuación, se calculará el error de predicción del modelo como el porcentaje de observaciones mal clasificadas en la matriz de confusión. Siendo su valor 16.09977.

Seguidamente, se pasa a calcular la curva ROC, es una herramienta gráfica que permite evaluar el modelo de clasificación y compararlo con otros futuros modelos.

Figura 8: Curva ROC Modelo 1



Fuente 9: Elaboración propia

Como se puede apreciar se encuentran la tasa de verdaderos positivos en el eje Y, y la tasa de falsos positivos en el eje x. Para interpretar el significado de la curva ROC se debe tener en cuenta que cuanto más se separe la curva de la línea diagonal imaginaria, mejor es el rendimiento (Roy-García et al.,2023). Por tanto, cuanto mayor sea el área que hay debajo de la curva “Area Under the Curve (AUC)”, mejor. En este caso, el AUC toma un valor de 0,81, teniendo en cuenta que si tomase valor 1 sería un modelo perfecto (Sanahuja & de Pol,2021).

También se calcularán las métricas de precisión, especificidad y sensibilidad. Este primer modelo tiene una precisión de 0,83. Hace referencia a que el modelo tiene una probabilidad de acertar en sus predicciones del 83%. Toma valores de especificidad y sensibilidad de 0,88 y 0,58 respectivamente. Con ello se puede afirmar que el modelo clasifica mejor las observaciones que realmente son negativas como negativas, el 88%, que las que realmente son positivas como positivas, el 58%.

Esta afirmación es una problemática a tener en cuenta, dado que como se ha mencionado anteriormente, el número de falsos negativos es más alto para el modelo. Esto significa que al modelo le cuesta más predecir bien la gente que va a renunciar, dando lugar a que empleados que tienen en mente abandonar la empresa de manera voluntaria sean clasificados como que no, lo que no nos permitiría tomar medidas al respecto para frenar esa renuncia voluntaria al no poder identificar el caso.

Asimismo, se ha calculado el valor de la métrica F1-score. Se considera que esta métrica podía tener valor para manejar las clases desbalanceadas, dado que se aprecia que una clase es mucho más frecuente que la otra (Lipton, 2014).

Esta medida tiene en cuenta tanto la precisión como la exhaustividad. Es decir, la primera mide la exactitud de las predicciones positivas del modelo y la segunda métrica establece la capacidad del modelo de identificar las instancias positivas. Se ha obtenido un valor de 0,6, lo que indica que el rendimiento de nuestro modelo es moderado. En un futuro podrían implementarse técnicas de re-muestreo para intentar mejorar esta métrica (Chicco & Jurman, 2020).

MODELO 2:

Para el segundo modelo, en lugar de incluir todas las variables como en el modelo anterior, se opta por utilizar el método "stepwise" como técnica de selección de variables. Esta elección se fundamenta en la búsqueda de un equilibrio entre rendimiento y simplicidad en el modelo final, comparado con otros métodos de selección de variables como "forward" y "backward".

El método "stepwise" nos permite determinar qué variables predictoras incluir en el Modelo2 de manera automatizada. Su objetivo es encontrar un modelo óptimo al considerar tanto el rendimiento predictivo como la simplicidad del modelo. Al utilizar la función "step" con el parámetro de dirección establecido en "both", el algoritmo realiza tanto adiciones como eliminaciones de variables en cada paso de su evaluación.

Por otro lado, el método "forward" se centra únicamente en la adición de variables al modelo de manera progresiva, evaluando en cada paso qué variable añadirá el mayor beneficio en términos de mejora en el rendimiento predictivo. Mientras tanto, el método

"backward" comienza con un modelo que incluye todas las variables y, en cada paso, elimina la variable que menos contribuye al rendimiento predictivo.

Comparado con "forward" y "backward", el enfoque "stepwise" presenta varias ventajas. En primer lugar, es computacionalmente más eficiente que el método exhaustivo, ya que evalúa un número menor de combinaciones posibles de variables. Esto lo hace especialmente útil en conjuntos de datos grandes o con un gran número de variables predictoras. Además, "stepwise" tiende a producir modelos más simples y fáciles de interpretar al eliminar variables redundantes o poco informativas. Esto ayuda a evitar el sobreajuste del modelo, donde se incluyen demasiadas variables irrelevantes que pueden afectar negativamente su capacidad predictiva en datos nuevos.

Figura 9: Salida función "glm" Modelo 2

```

> summary(step_model)

Call:
glm(formula = Attrition ~ OverTime + TotalworkingYears + MaritalStatus_Single +
  EnvironmentSatisfaction + JobSatisfaction + BusinessTravel_Travel_Frequently +
  JobInvolvement + NumCompaniesworked + YearsSinceLastPromotion +
  Department_ResearchandDevelopment + JobRole_LaboratoryTechnician +
  JobRole_SalesExecutive + EducationField_Medical + Relationshipsatisfaction +
  YearsInCurrentRole + Age + BusinessTravel_Travel_Rarely,
  family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5717  -0.4879  -0.2929  -0.1178   3.9861

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.01911    0.83951   3.596 0.000323 ***
OverTime        1.85736    0.22263   8.343 < 2e-16 ***
TotalworkingYears
-0.11012        0.02635  -4.178 2.94e-05 ***
MaritalStatus_Single
 1.13804        0.21442   5.308 1.11e-07 ***
EnvironmentSatisfaction
-0.43881        0.09482  -4.628 3.70e-06 ***
JobSatisfaction
-0.44154        0.09557  -4.620 3.84e-06 ***
BusinessTravel_Travel_Frequently
 1.62168        0.45076   3.598 0.000321 ***
JobInvolvement
-0.49275        0.14283  -3.450 0.000561 ***
NumCompaniesworked
 0.18405        0.04309   4.271 1.94e-05 ***
YearsSinceLastPromotion
 0.19317        0.04449   4.341 1.42e-05 ***
Department_ResearchandDevelopment
-1.45163        0.31516  -4.606 4.10e-06 ***
JobRole_LaboratoryTechnician
 1.07522        0.27710   3.880 0.000104 ***
JobRole_SalesExecutive
-0.77432        0.34133  -2.269 0.023294 *
EducationField_Medical
-0.60251        0.23956  -2.515 0.011902 *
Relationshipsatisfaction
-0.20955        0.09590  -2.185 0.028876 *
YearsInCurrentRole
-0.09652        0.04319  -2.235 0.025424 *
Age
-0.02737        0.01585  -1.727 0.084203 .
BusinessTravel_Travel_Rarely
 0.65267        0.42185   1.547 0.121825

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 906.03  on 1028  degrees of freedom
Residual deviance: 622.23  on 1011  degrees of freedom
AIC: 658.23

```

Fuente 10: Elaboración propia

Como resultado obtuvimos un Modelo2 en el cual se incluyeron las variables que aparecen en la figura 9, estas son horas extra, número total de años en el mundo laboral, Estado civil-soltero, satisfacción con el entorno, viajes de negocios-frecuentemente,

grado de compromiso laboral, número de empresas en las que ha trabajado, años desde la última promoción, Departamento- Investigación y desarrollo, Rol laboral_Técnico de laboratorio, Rol laboral-Ejecutivo de ventas, Campo de educación- Medico, años en el rol actual, edad, viajes de negocios – raramente.

Sin embargo, como ya se explicó en el desarrollo del Modelo1, las variables más influyentes en los resultados de este segundo modelo son aquellas cuyo p-valor es menor.

Se obtuvo la matriz de confusión, para comparar valores predichos frente valores reales. Como ya explicamos anteriormente en la matriz de Modelo1, podemos observar que el modelo clasificó a 40 empleados como que pretendían abandonar su puesto de trabajo de manera voluntaria cuando realmente así era. Los verdaderos negativos toman un valor de 329 empleados que el modelo identificó de manera correcta como empleados que no pretendían renunciar de manera voluntaria cuando efectivamente no renunciaron.

Además, el modelo determinó 40 casos de falsos positivos, siendo estos empleados clasificados incorrectamente como “renuncia voluntaria” cuando en realidad no renunciaron. Finalmente, 32 empleados fueron asignados a la clase de falsos negativos. Estos serían los empleados que el modelo clasificó incorrectamente como “No renuncia voluntaria” y en realidad si abandonaron su puesto de trabajo. Este es el error más perjudicial.

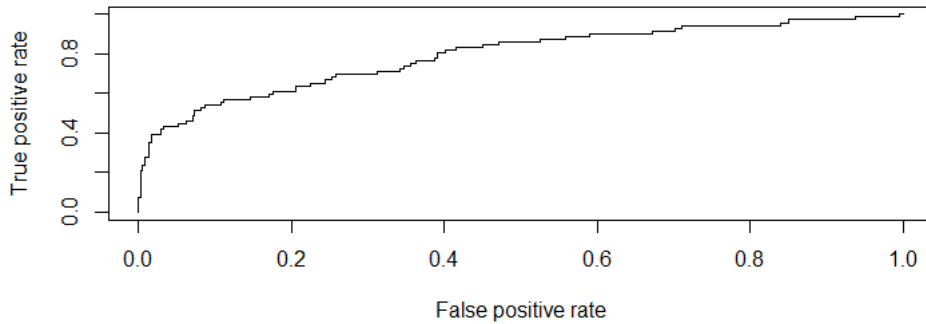
Figura 10: Matriz de confusión Modelo 2

```
> matrizConfusion
      REAL
PREDICTED  0  1
FALSE     329 32
TRUE      40 40
```

Fuente 11: Elaboración propia

Una vez realizado el análisis de la matriz se puede calcular su error, con valor de 16.32653, siendo algunas décimas mayores que en el Modelo1. Asimismo, en la figura 11, podemos apreciar la curva ROC, cuyo valor del área bajo la curva toma un valor de 0,79, unas décimas más bajas que el Modelo1.

Figura 11: Curva ROC Modelo 2



Fuente 12: Elaboración propia

Una precisión, sensibilidad y especificidad de 0,83, 0,55 y 0,89 respectivamente. Podemos interpretar estos datos como que el segundo modelo tiene una probabilidad del 83% de acierto en sus predicciones. Identifica de manera correcta a los verdaderos positivos el 55% de los casos y el 89% identifica bien a los verdaderos negativos. En estos aspectos, tanto el Modelo1 como el Modelo2 toman valores muy similares. Y por tanto ambos presentan el problema que hemos enunciado anteriormente. El modelo tiene una tendencia a tener un alto número de falsos negativos. Esto implica que el modelo tiene dificultades para predecir correctamente cuándo los empleados están considerando renunciar. Como resultado, algunos empleados que están pensando en dejar la empresa podrían ser clasificados erróneamente como trabajadores que no pretenden abandonar la empresa, lo que impediría tomar acciones preventivas para evitar su renuncia, ya que no se podrían identificar a trabajadores poco satisfechos que tienen intención de abandonar su puesto de trabajo. Además, la métrica F1-score también es muy similar a la obtenida en el modelo anterior.

A modo de conclusión se establece que son dos modelos muy similares en términos de error, precisión, sensibilidad y especificidad. No obstante, el Modelo1 tiene un valor del criterio de información de Akaike (AIC) de 682,45 y el AIC de Modelo2 toma valor 658,23. El criterio de información de Akaike tiene un papel de equilibrio entre bondad de ajuste y complejidad del modelo. Esto quiere decir que cuanto menor es el valor del AIC nuestro modelo tiene un mejor balance entre cómo se ajusta a los datos y lo complejo que es.

Aunque la diferencia es bastante pequeña entre los modelos y se debería seguir investigando otros criterios para afirmar esto de forma definitiva, se considera mejor el segundo modelo, dado que, posee un valor de AIC menor (Martínez et al.,2009).

No obstante, se es consciente de que el modelo predictivo utilizando el algoritmo de regresión logística no alcanza el nivel de sensibilidad esperado. Como el hecho de que el modelo prediga únicamente con un 58% de acierto los casos en los que un empleado vaya a abandonar una empresa y realmente lo haga. Lo que se traduce en que un alto porcentaje de empleados que tienen en mente llevar a cabo una renuncia voluntaria no serán identificados por el modelo. Por eso mismo se decide realizar un modelo predictivo utilizando el algoritmo de random forest, con el fin de obtener mejores resultados.

DESCRIPCIÓN DEL ALGORITMO RANDOM FOREST

El algoritmo Random Forest es una técnica de conjunto de modelos que se basa en la técnica de bagging. En este enfoque, se crean múltiples modelos de árboles de decisión, cada uno entrenado con una muestra aleatoria con reemplazo de los datos originales. Para cada árbol, se selecciona un subconjunto aleatorio de variables en cada división del árbol, siendo el número de variables seleccionadas un hiperparámetro clave del algoritmo.

Una vez cada uno de los distintos árboles ha generado sus predicciones a cerca de “¿A qué clase pertenece la observación?” Se hace una votación entre esos resultados, seleccionándose finalmente el resultado de la mayoría. Este algoritmo nos permite mitigar la alta variabilidad inherente a cada árbol de decisión individual, lo que resulta en predicciones más estables y precisas. Asimismo, cuanto menos correlacionados estén los árboles entre sí, mejor será el rendimiento del modelo obtenido (DataScientest, 2022).

Primeramente, se empieza por realizar el tratamiento indicado para la base de datos, se siguen exactamente los mismos pasos que en los modelos anteriores, pero no se llevarán a cabo el tratamiento de variables “One-hot encoding”. Además, en esta ocasión se convertirá la variable objetivo “Attrition” que representa si un empleado renuncia voluntariamente o no, en tipo factor.

A continuación, realizamos la partición del conjunto de datos en entrenamiento (70%) y validación (30%) al igual se hizo en el caso anterior y se dispone a entrenar el

modelo utilizando la función “RandomForest” en conjunto de datos de entrenamiento. Más adelante se realizan las predicciones con el ModeloRandomForest sobre el conjunto de validación, obteniendo la matriz confusión. La intención al realizar las predicciones haciendo uso del conjunto de “test”, es poder evaluar el rendimiento de nuestro último modelo.

Figura 12: Matriz de confusión Modelo 3

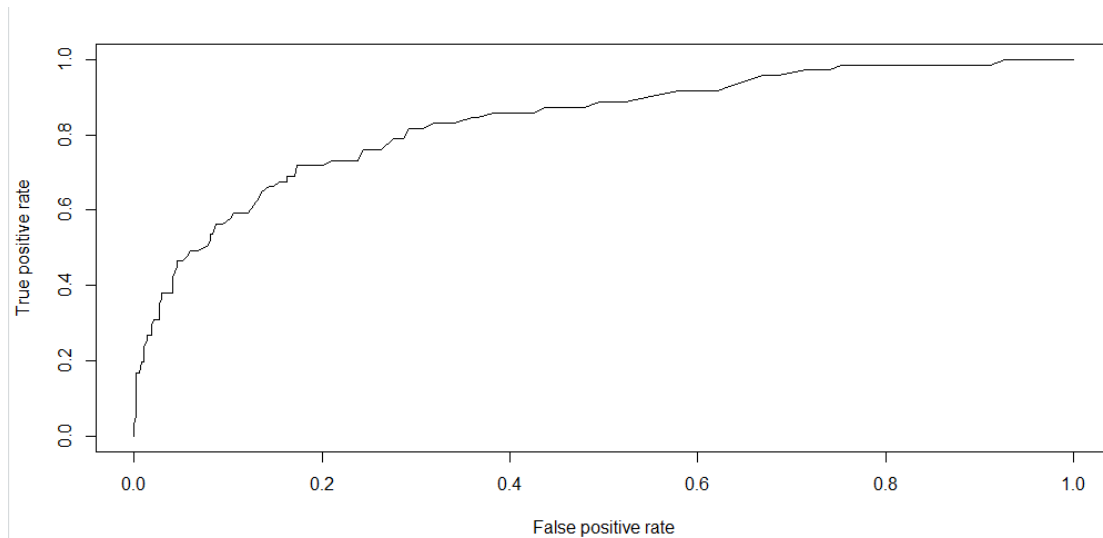
	Predicho	
Real	0	1
0	365	4
1	55	16

Fuente 13: Elaboración propia

Haciendo uso de los resultados proporcionados por la matriz se pudo establecer que el modelo tenía una precisión del 86%, una sensibilidad de 0,8 y especificidad de 0,86. Esto pone de manifiesto que el modelo generado utilizando el algoritmo de RandomForest tiene mejor rendimiento que los obtenidos a través del algoritmo de regresión logística. Se observa que, este último modelo consigue solventar el bajo nivel de sensibilidad que demostraban los otros. Por tanto, con este nuevo modelo se es capaz de identificar en el 80% de los casos a empleados que tengan pensado abandonar su puesto de trabajo, y realmente así sea. Por tanto, serán mucho menos los empleados que no se puedan identificar con el modelo como “renuncia voluntaria” correctamente. De esta manera el número de trabajadores sobre los que no se puedan tomar medidas para prevenir el abandono del puesto de trabajo será mucho menor. Resolviendo así, en gran medida, el problema que teníamos con los modelos generados mediante regresión logística.

Por último, se calcula la curva ROC del modelo RandomForest y El valor del área bajo la curva (AUC), que toma valor de 0,83, siendo 1 el valor que haría referencia a un modelo perfecto.

Figura 13: Curva ROC Modelo 3



Fuente 14: Elaboración propia

Tabla 2: Resumen modelos

	Precisión	Sensibilidad	Especificidad	AUC
Regresiónlog 1	0,83	0,88	0,58	0,81
Regresiónlog 2	0,83	0,89	0,55	0,79
RandomForest	0,86	0,86	0,8	0,83

Fuente 15: Elaboración propia

A modo de conclusión, se presenta una tabla resumen para ver una comparativa de los tres modelos ejecutados. Salta a la vista que el modelo RandomForest es el que mejor realiza las predicciones, puesto que presenta el nivel más alto de precisión y predice correctamente de manera equitativa tanto los verdaderos positivos como los verdaderos negativos. De igual manera, posee el valor más elevado de “área bajo la curva (AUC)”.

Finalmente, se presentan algunas consideraciones por las cuales pensamos que nuestros modelos de regresión logística han podido obtener peores resultados. Primeramente, el hecho de haber realizado una codificación especial siguiendo el método de “one-hot encoding” para variables categóricas puede afectar en el resultado de las predicciones. Asimismo, los modelos de regresión logística son más sensibles al factor de multicolinealidad. Los modelos Random Forest son más robustos ante este tipo de situaciones. Por último, la regresión logística presenta mayor tendencia al sobreajuste.

Por el contrario, los modelos de árboles de decisión tienden a reducir la varianza del modelo.

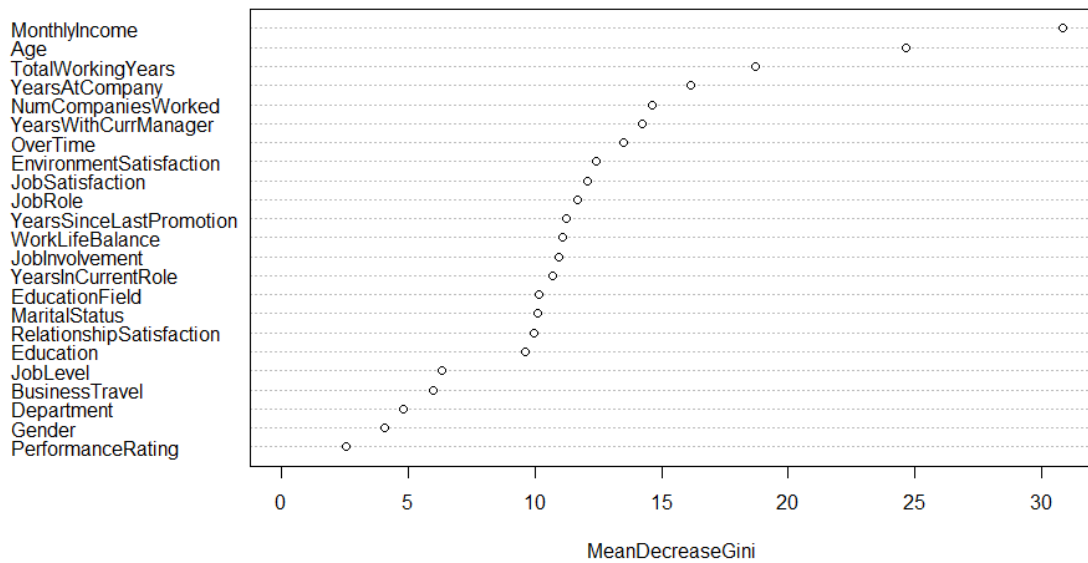
4 ANÁLISIS DE RESULTADOS

4.1 Interpretación de resultados

Una vez terminado el trabajo de investigación resulta fundamental llevar a cabo una interpretación clara de los resultados obtenidos. Durante la investigación se ha llevado a cabo la elaboración de tres modelos predictivos. Dos de ellos a través del algoritmo de regresión logística y el tercero mediante el uso del algoritmo de “Random Forest”. Sin embargo, como se ilustra en la tabla 2, donde aparece una tabla resumen de los resultados obtenidos en los tres modelos, aquel que ofrece mejores resultados y mayor rendimiento es el último modelo obtenido utilizando “RandomForest”. Consecuentemente, se interpretarán los resultados que provienen de este tercer modelo. En primer lugar, se pretende identificar cuáles son las variables independientes más relevantes en el tercer modelo predictivo. De esta manera se sabrá cuáles son los factores que más influyen a un trabajador a la hora de tomar la decisión de abandonar su puesto en la empresa para la cual trabaja o, por el contrario, mantenerlo.

Asimismo, resulta evidente que las variables que toman mayor importancia en este tercer modelo de “Random Forest” son distintas a las que se consideraron más relevantes en los modelos de regresión logística, lo cual resulta coherente dado a la disparidad de los resultados obtenidos con ambos modelos.

Figura 14: Variables modelo "Random Forest"



Fuente 16: Elaboración propia

Este gráfico representa la importancia relativa de las variables utilizadas en nuestro modelo “Random Forest”. Es una herramienta útil para identificar qué variables tienen un mayor impacto en la predicción de si un empleado continuará en la empresa o no (Cutler et al. ,2022).

En el eje y, se muestran los nombres de las distintas variables independientes utilizadas en la predicción, mientras que en el eje x se presenta la disminución del índice de Gini promedio cuando se prescinde de cada una de las variables (MeanDecreaseGini). Este índice es una medida de la impureza en los nodos de cada árbol de decisión. Cuanto mayor sea la disminución de este índice para una variable, mayor será su importancia para el modelo (GeeksforGeeks, 2022). En otras palabras, las variables con valores más altos en el eje x tienen un mayor impacto en la predicción de la retención de empleados.

A continuación, se presentan las diez variables que más afectan a los trabajadores a la hora de renunciar voluntariamente a un trabajo ordenadas de manera descendente según nuestro último modelo:

- MonthlyIncome: La variable de salario mensual es la más relevante a la hora de tomar la decisión de renunciar voluntariamente por parte de un empleado. Resulta evidente la importancia del salario en nuestro puesto de trabajo, aunque puede variar según la situación y las preferencias individuales de cada trabajador

- Age: La edad es otro factor muy importante a la hora de abandonar el puesto de trabajo. Se hace patente que no presentan la misma situación personal, ni tienen acceso a las mismas oportunidades una persona que acaba de iniciar su andadura en el mundo profesional que alguien que está cerca de finalizarla.
- TotalWorkingYears: El número total de años trabajando resulta significativo como hemos mencionado en el punto anterior.
- YearsAtCompany: La variable de número de años en la empresa tiene cierto efecto a la hora de decidir abandonarla.
- NumCompaniesWorked: El número de compañías en las que ha trabajado el empleado. Pueden existir distintos perfiles de trabajadores más nómadas o sedentarios.
- YearsWithCurrentManager: La variable que representa el número de años que el empleado lleva con el mismo “manager”.
- OverTime: El número de horas extra realizadas por el trabajador, también es una variable relevante en cuanto a tomar la decisión de renuncia voluntaria
- EnvironmentSatisfaction: La variable que representa la satisfacción con el ambiente de trabajo indudablemente influye si se pretende abandonar una empresa.
- JobSatisfaction: El grado de satisfacción con la empresa en general influye en la opinión de un trabajador sobre la misma.
- JobRole: El puesto que ocupe el empleado también es un factor fundamental en relación con la decisión de llevar a cabo una renuncia voluntaria

Aunque los modelos predictivos como la regresión logística y Random Forest han proporcionado valiosos” insights “, tienen algunas limitaciones.

Los datos pueden no representar todas las situaciones, especialmente si contienen sesgos hacia ciertas demografías o roles específicos, lo que afecta la generalización del modelo. Además, las condiciones laborales y económicas cambian con el tiempo, y un modelo basado en datos históricos puede no capturar adecuadamente estas dinámicas, afectando su precisión futura.

Por eso mismo, en un futuro podría mejorarse el modelo considerando varias estrategias. Por ejemplo, diversificar los datos incluyendo diferentes industrias, regiones

y perfiles de empleados, mejorando la representatividad del modelo; utilizar algoritmos más avanzados como redes neuronales profundas o implementar sistemas de aprendizaje continuo permitirá actualizar el modelo con datos nuevos y relevantes, manteniéndolo alineado con las condiciones actuales del mercado y del lugar de trabajo.

Estas mejoras desarrollarán modelos más precisos y útiles.

A modo resumen, en el estudio, se elaboraron tres modelos predictivos: dos con regresión logística y uno con “Random Forest”. Este último mostró mejores resultados y mayor rendimiento. Se centró la atención en las variables independientes más relevantes del modelo “Random Forest”. Las variables más importantes, según la disminución del índice de Gini, son: salario mensual, edad, años totales de trabajo, años en la empresa, número de compañías trabajadas, años con el mismo jefe, horas extra, satisfacción con el ambiente, satisfacción laboral y puesto ocupado por el empleado.

4.5 Visualizaciones

A continuación, se realizarán una serie de gráficos para comparar cada una de las variables que se han mencionado anteriormente con la variable objetivo. Pudiendo observar así, la relación que se presenta entre ellas.

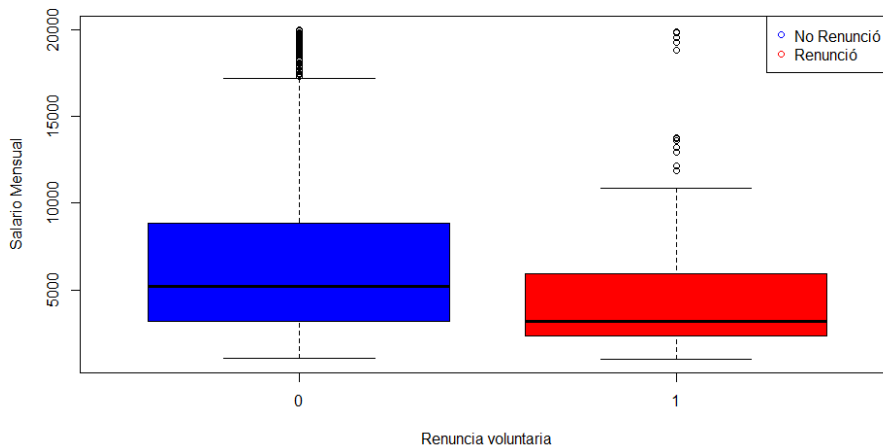
Para las primeras visualizaciones se utilizará el diagrama de bigotes. La mediana es una medida menos sensible a valores atípicos más conocidos como “outliers” que la media. Por eso, esta medida servirá de guía, que viene representada como una línea horizontal negra oscura en vez de por la media. La caja en sí misma representa el rango intercuartílico, desde el primer cuartil hasta el tercero. Los datos por debajo del primer cuartil representan el 25% de los mismos. Mientras que aquellos que se encuentran por debajo del tercer cuartil, representa en 75% de los datos. A continuación, aparecen los “bigotes”, que están representados por las líneas que salen de la caja central hasta los valores mínimos o máximos sin tener en cuenta los valores atípicos.

- Salario mensual Vs Renuncia voluntaria

Este diagrama representa la distribución de la variable “salario mensual” en función de si la variable “renuncia voluntaria” toma valor positivo o negativo.

Gracias a este informe somos capaces de comparar fácilmente las distribuciones entre los grupos. La mediana representa el valor central de un conjunto de datos ordenados y como ya se ha mencionado anteriormente es una medida menos sensible a “outliers”. Se aprecia de manera clara que la mediana del grupo de trabajadores donde la variable “renuncia voluntaria” toma valor 1 tiene un valor menor. Esta afirmación, pone de manifiesto que aquellos empleados que toman la decisión de renunciar tienen un salario menor que aquellos que permanecen en su puesto de trabajo.

Figura 15: Salario mensual Vs Renuncia voluntaria

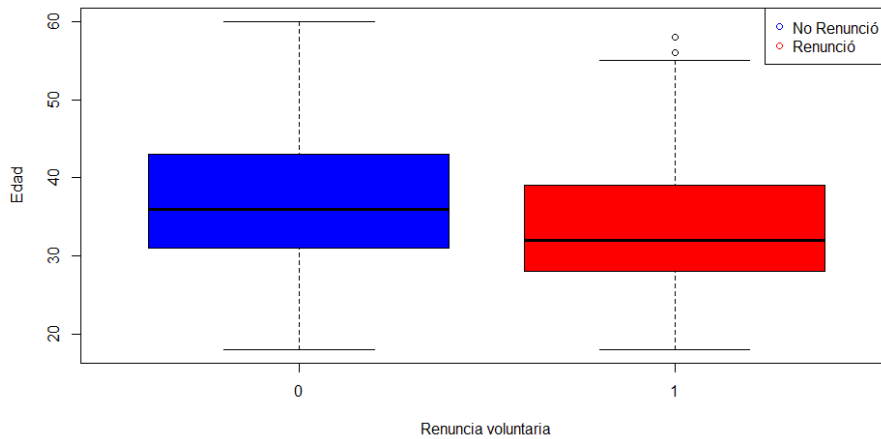


Fuente 17: Elaboración propia

- Edad Vs Renuncia voluntaria

Este diagrama de cajas ejemplifica que por norma general la edad de los trabajadores que abandonan su puesto de trabajo voluntariamente es menor que la de aquellos que permanecen. Esto puede ocurrir por muchas razones como que los trabajadores más jóvenes todavía están buscando una empresa en la que encajar y que se ajusten a sus preferencias entre ellas.

Figura 16: Edad Vs Renuncia voluntaria

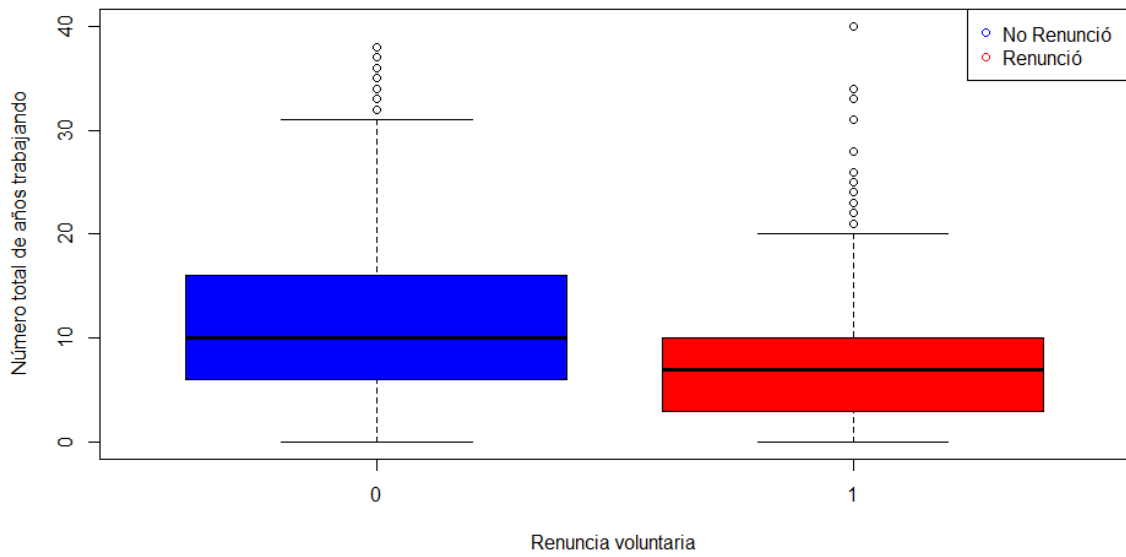


Fuente 18: Elaboración propia

- Número total de años trabajando Vs Renuncia voluntaria

En esta visualización se puede apreciar como el valor de la mediana de la variable “número total de años trabajando” para el grupo de empleados que renuncia voluntariamente toma un valor más pequeño que para aquellos que permanecen en la empresa. Esta afirmación puede estar correlacionada con la relación mencionada en el gráfico anterior. Es decir, que cuanto más joven es un trabajador, menos años lleva en el mundo laboral, y por tanto más probabilidades hay de que el empleado quiera explorar otros puestos de trabajo u otras empresas. Asimismo, también pueden contribuir otros factores como los contratos de prácticas, las expectativas no cumplidas, problemas de adaptación, etc.

Figura 17: Número total de años trabajando Vs Renuncia voluntaria

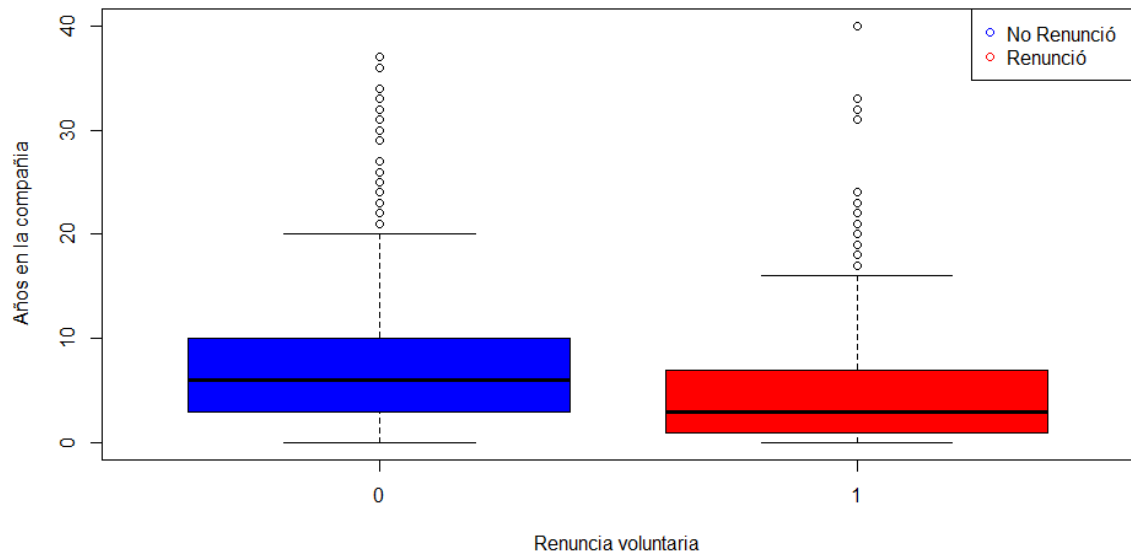


Fuente 19: Elaboración propia

- Años en la compañía Vs Renuncia voluntaria

Este diagrama de cajas compara las variables “Años en la compañía” y “Renuncia voluntaria”. Esta visualización pone de manifiesto que los trabajadores que renuncian de manera voluntaria a su puesto de trabajo llevan menos años en la compañía que aquellos que permanecen en su puesto. Esto puede deberse a factores como falta de arraigo y conexión con la cultura de la empresa en la que trabajan.

Figura 18: Años en la compañía Vs Renuncia voluntaria



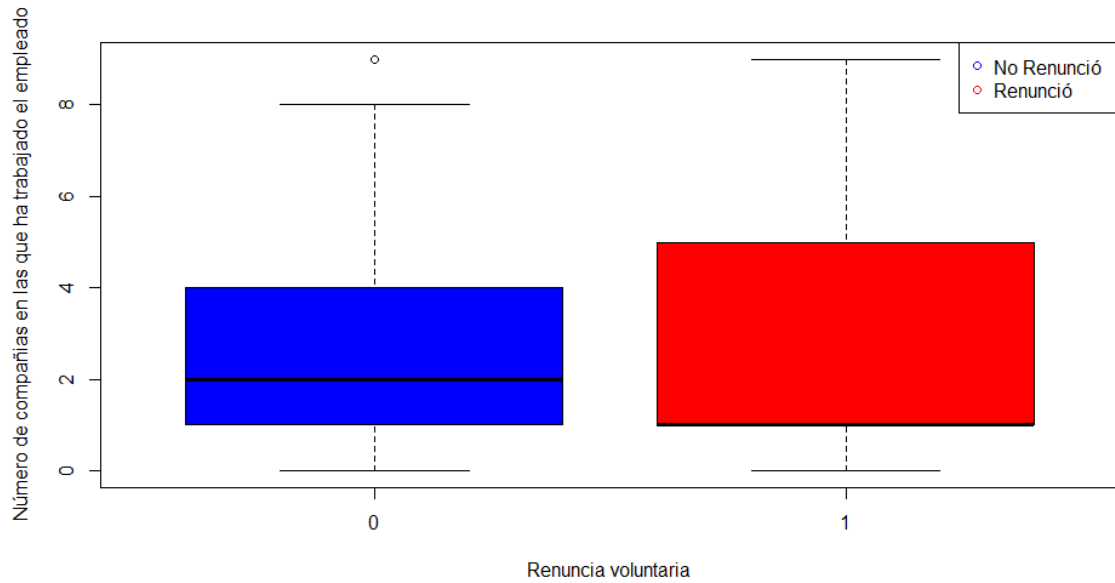
Fuente 20: Elaboración propia

- Número de compañías en las que ha trabajado el empleado VS renuncia voluntaria

En este gráfico comparativo se aprecia que la mediana del número de compañías en las que ha trabajado un empleado que decide abandonar su puesto de trabajo de manera voluntaria, es menor que la de aquellos trabajadores que no renuncian. Esta afirmación presenta coherencia con las aseveraciones anteriores. Dado que cuanto más joven es un trabajador, seguramente haya pasado por menos empresas y tenga más curiosidad por explorar otras opciones hacia las que orientar su carrera profesional

Igualmente, cabe resaltar el hecho de que la mediana del grupo en el que la variable “Renuncia voluntaria” toma valor 1 coincide con el primer cuartil. Esto significa que al menos un 25% de los datos se encuentran por debajo de la mediana. Lo cual indica que la distribución de datos está más sesgada hacia valores inferiores.

Figura 19: Número de compañías en las que ha trabajado el empleado Vs Renuncia voluntaria

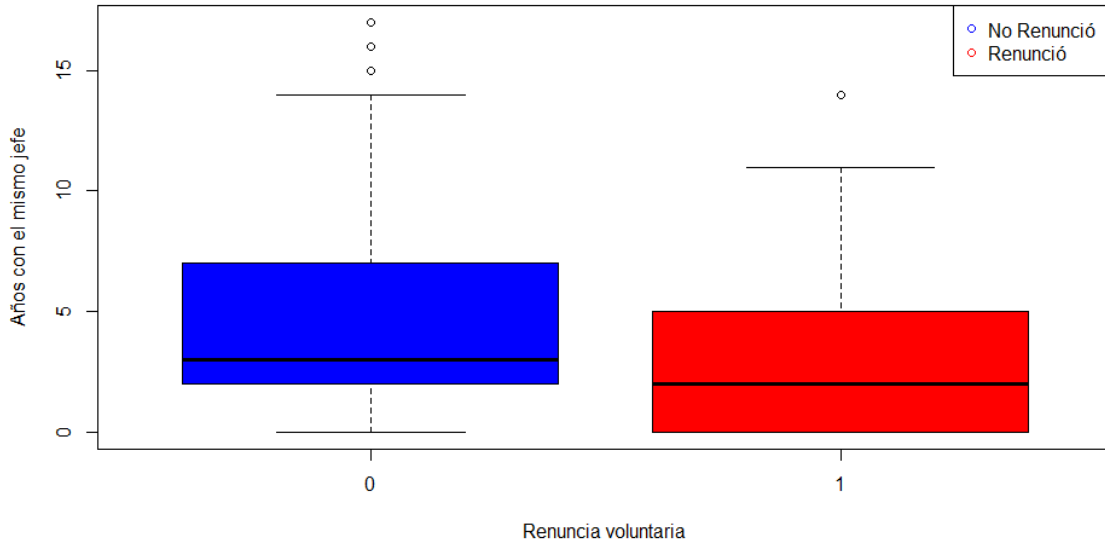


Fuente 21: Elaboración propia

- Años con el mismo jefe Vs Renuncia voluntaria

En esta visualización se pone de manifiesto cómo el grupo de trabajadores que deciden renunciar voluntariamente llevan menos años con el mismo jefe que aquellos que deciden mantener su puesto de trabajo. Se pueden apreciar ciertos “outliers”, pero por norma general esta es la premisa que tiene lugar. Puede deberse a factores como tener menos estabilidad y conexión en su entorno laboral. Adicionalmente el estar cambiando continuamente de supervisor puede dificultar la creación de una relación de confianza o generar incertidumbre.

Figura 20: Años con el mismo jefe Vs Renuncia voluntaria

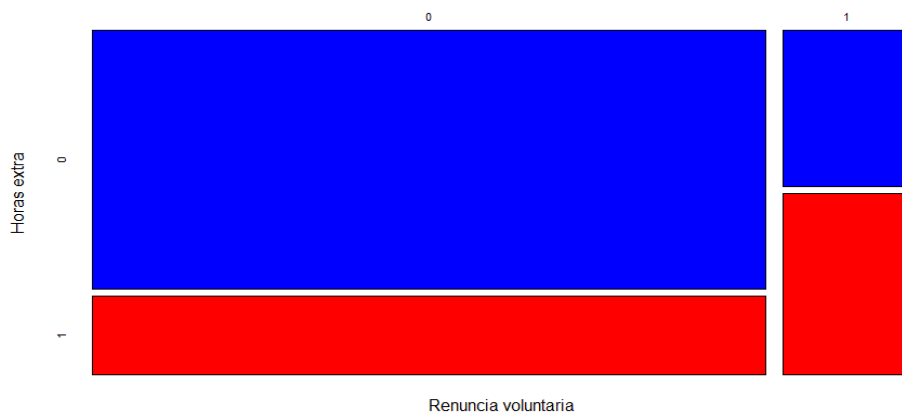


Fuente 22: Elaboración propia

- Horas extra Vs Renuncia voluntaria

La siguiente visualización muestra una comparativa de las variables “Horas extra” y “Renuncia voluntaria”. Salta a la vista como el cuadrante en el que la variable renuncia voluntaria toma valor uno, la variable horas extra también toma valor uno en mayor proporción que valor cero. Resulta evidente que aquellos empleados que realicen una mayor cantidad de horas extra estarán más descontentos en su puesto de trabajo y por tanto tendrán más propensión al abandono voluntario de su puesto de trabajo.

Figura 21: Horas extra Vs Renuncia voluntaria

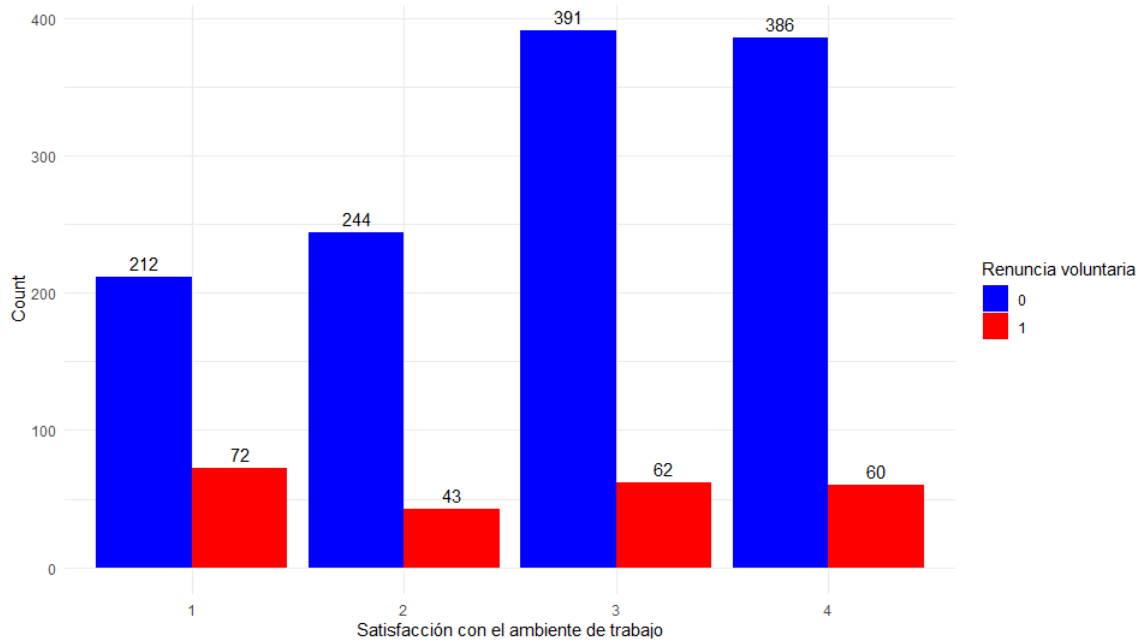


Fuente 23: Elaboración propia

- Satisfacción con el ambiente de trabajo Vs Renuncia voluntaria

La siguiente visualización representa un gráfico de barras de los valores que toma la variable “Satisfacción con el ambiente de trabajo” para el valor cero y uno de la variable “Renuncia voluntaria”. No obstante, para poder hacer una comparativa hemos calculado el porcentaje de renuncia voluntaria que hay en cada nivel de satisfacción con el ambiente de trabajo. La variable de satisfacción hace referencia al entorno general de trabajo de un empleado, y los valores van del uno al cuatro, ordenados de menor a mayor respectivamente. Cuando el valor de satisfacción toma valor “Bajo”, la variable “Renuncia voluntaria” toma valor uno, en un 29% de los casos. Si toma valor “Medio” un 15% de los empleados renuncia voluntariamente. Si la satisfacción con el ambiente de trabajo toma valor “Alta” un 14% de los trabajadores renuncian de manera voluntaria. Por último, si la satisfacción es “Muy alta” únicamente un 13% de los empleados renuncia de manera voluntaria. Tras realizar el análisis de este doble gráfico de barras podemos afirmar que, a mayor grado de satisfacción con el ambiente de trabajo, más disminuye la renuncia voluntaria en la empresa. Parece coherente que los empleados que están satisfechos con su ambiente de trabajo disfruten de una mejor calidad de vida y estén más motivados con su trabajo. No obstante, la mayor diferencia está entre el escalón bajo y medio, dado que a partir del nivel medio de satisfacción los porcentajes de renuncia disminuye sutilmente.

Figura 22: Satisfacción con el ambiente de trabajo Vs Renuncia voluntaria

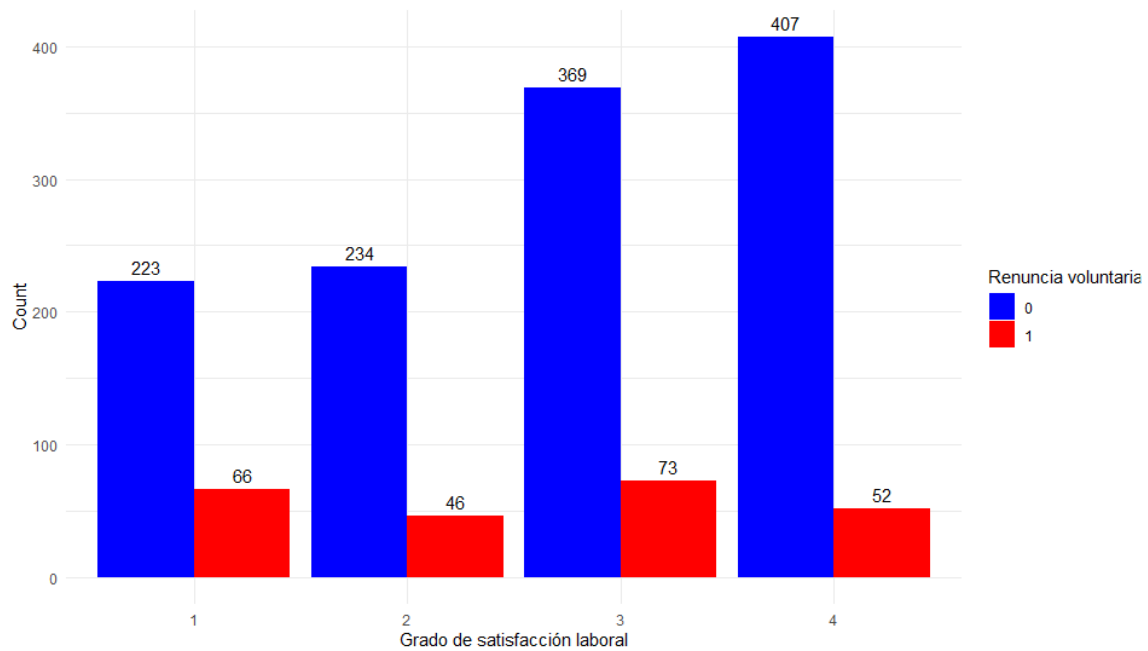


Fuente 24: Elaboración propia

- Satisfacción laboral Vs Renuncia voluntaria

Este gráfico de barras muestra la presencia de la variable “Renuncia voluntaria” tomando valores de uno o cero según el grado de satisfacción laboral. Al igual que en gráfico anterior la variable toma valores de uno a cero siendo uno el menor grado de satisfacción y cuatro el mayor. La diferencia entre esta variable de satisfacción y la anterior es que la variable anterior representa principalmente la satisfacción con la oficina, los compañeros, el trato de los superiores, etc. Y esta variable se centra más en la satisfacción directamente con el puesto de trabajo sin tener en cuenta la atmósfera de alrededor del empleado. Se sigue el mismo procedimiento que en la visualización anterior para obtener los porcentajes de renuncia voluntaria dentro de cada nivel de satisfacción. Si el nivel de satisfacción con el trabajo toma valor “Bajo” la renuncia voluntaria es de un 22%, si toma valor “Medio” de un 17%, si toma valor “Alto” de un 16% y si toma valor “Muy alto” la renuncia voluntaria disminuye a un 11%. Al igual que en la representación anterior según va incrementándose el nivel de satisfacción con el trabajo, se va disminuyendo el porcentaje de empleados que decide abandonar su puesto en la empresa. Se aprecia como el mayor salto de porcentajes esta entre los niveles “Bajo-Medio” y “Alto-Muy alto”.

Figura 23: Satisfacción laboral Vs Renuncia voluntaria

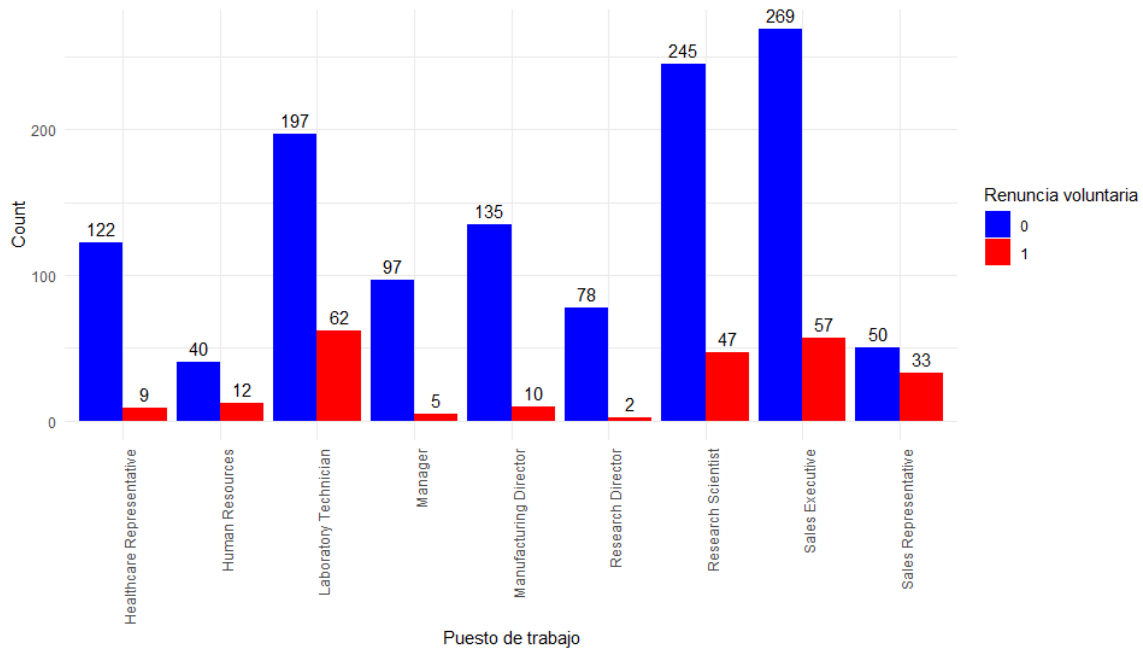


Fuente 25: Elaboración propia

- Puesto de trabajo Vs Renuncia voluntaria

Finalmente, se utiliza un gráfico de barras para realizar una comparativa entre los diferentes niveles de renuncia voluntaria según la jerarquía de algunos puestos en distintos sectores del mundo laboral. Se van a ir ordenando los puestos de mayor a menor según el porcentaje de trabajadores que renuncian voluntariamente. Asimismo, se establecen una serie de razones por las que creemos que los empleados de los diferentes puestos de trabajo pueden plantearse abandonar su empresa de manera voluntaria teniendo en cuenta el rol que desempeñan. Es importante resaltar que hay muchos empleos que no están representados en esta muestra que se deberían analizar en un futuro incrementando el tamaño de la misma.

Figura 24: Puesto de trabajo Vs Renuncia voluntaria



Fuente 26: Elaboración propia

En primer lugar, se tiene a los representantes de ventas con un 40% de trabajadores que renuncian voluntariamente. Esto puede deberse a diferentes casuísticas como pueden ser la alta presión por cumplir cuotas de ventas, el estrés o agotamiento y la incertidumbre de las comisiones pueden llevar a que el grado de rotación sea prácticamente de la mitad de sus trabajadores.

En segunda posición, el técnico de laboratorio, donde el porcentaje de renuncia voluntaria es del 24%. Los trabajadores pueden verse motivados a abandonar su puesto a causa de la exigencia implícita en los horarios de guardia. Por eso, la búsqueda de roles con mayor equilibrio entre vida laboral y personal puede generar mayor rotación en el sector.

A continuación, el sector de recursos humanos, donde el porcentaje de trabajadores que renuncian de manera voluntaria asciende al 23%. Pensamos que la rotación puede deberse a la discrepancia con la cultura organizacional de la empresa o la búsqueda de otras oportunidades de desarrollo profesional.

En cuarto lugar, se encuentran los ejecutivos de ventas, con un 17% de porcentaje de empleados que renuncian voluntariamente. Establecemos que las razones son similares al caso de los representantes de ventas.

Adicionalmente, los científicos de investigación son los siguientes en la lista con un porcentaje del 16% de trabajadores que abandonan su puesto de trabajo por voluntad propia. Pensamos que esta rotación puede deberse a que se presenten nuevos proyectos de investigación más interesantes o avanzados.

Más adelante, aparecen los directores de manufactura y a los representantes de atención médica con un 6% de empleados que renuncian de manera voluntaria. Los primeros se encargan de supervisar todas las etapas del proceso productivo y se enfrentan a los desafíos de la gestión de operaciones, mejora de eficiencia y reducción de costes, lo que pueden generar estrés y presión en sus empleados. Por otro lado, la rotación en representantes de atención médica puede deberse a la exigencia a la hora de alcanzar objetivos de ventas.

En penúltimo lugar, aparece la posición de gerente con un porcentaje de renuncia voluntaria del 4,9%. Al ser un puesto más avanzado dentro de la carrera profesional, los trabajadores tienen mayor control sobre el estrés y gestión de emociones, gracias a su recorrido y competencia. Además, han recorrido un largo camino para alcanzar esa posición, por tanto, el número de renuncias voluntarias es menor. No obstante, el que los gerentes no se sientan respaldados por la empresa, o la aparición de otras oportunidades que les permitan desarrollarse más profesionalmente pueden ser causas que motiven el abandono voluntario.

Por último, se encuentran los directores de investigación con un 2,5% de porcentaje de renuncia voluntaria. Se piensa que la rotación puede ser menor debido a la experiencia y la especialización requeridas para este rol como hemos comentado anteriormente con los gerentes.

A modo resumen, se han utilizado estas visualizaciones para poner de manifiesto cómo las variables que han sido identificadas como más relevantes a la hora de tomar la decisión de abandonar un puesto de trabajo voluntariamente por el modelo “RandomForest”, se relacionan con la variable objetivo “Renuncia voluntaria”. De esta

manera, los gráficos anteriores ayudan a identificar patrones que tienden a identificar a aquellos empleados que tienen en mente abandonar su puesto de trabajo.

Una vez se han generado ciertas evidencias haciendo uso de las visualizaciones obtenidas, se dispondrán a contrastar las conclusiones obtenidas, con la literatura ya existente que gira en torno al tema de estudio.

5 DISCUSIÓN

5.1 Literatura existente

En esta sección se dedicará a contrastar las conclusiones obtenidas en el análisis de datos con la literatura ya existente. Manifestando la veracidad y congruencia de los hallazgos.

Se comentarán aquellas que se consideren más relevantes o sobre las que haya estudios más interesantes ya realizados.

En primer lugar, se hace referencia a estudios que ponen de manifiesto cómo el salario mensual tiene consecuencias en la rotación existente en una empresa. Según un trabajo de investigación realizado por la universidad de lima los empleados que consideran que su salario es insuficiente pueden sentirse motivados a buscar o aceptar ofertas más atractivas en otros lugares (Salazar,2022).

El percibir que la compensación económica recibida no permanece en sintonía con los deberes, conocimientos, experiencia y formación del empleado puede generar una sensación de frustración y llevar al trabajador a sentirse infravalorado. Desencadenando una percepción de rechazo por parte del empleado hacía la empresa (Torres,2023)

Por ende, ofrecer una remuneración competitiva puede desempeñar un papel crucial en la retención del talento y en la disminución de la rotación de empleados

Asimismo, el informe anual elaborado por la empresa líder en Recursos Humanos en España “Randstad”, publica un estudio que establece que la compensación mensual continúa siendo por tercer año consecutivo el factor más relevante a la hora de seleccionar una empresa por parte de los trabajadores. Concretamente su estudio determina que el 63% de los profesionales españoles sitúan la cuantía del salario mensual en el top 5 de sus preferencias al buscar empleo (Randstad, 2016).

Seguidamente, se centran en la relación establecida por el estudio entre la edad del empleado y tomar la decisión de abandonar la empresa de manera voluntaria. Tras llevar a cabo un proceso de investigación exhaustivo de la literatura existente se han encontrado varios estudios que avalan la idea de que los jóvenes tienden a renunciar voluntariamente a sus empleos más que las personas de edad más elevada. Según un

informe de LinkedIn publicado por la vanguardia siete de cada diez trabajadores entre 18 y 24 años estarán dispuestos a abandonar su puesto actual por otro que les convenga más. Según el informe las prioridades de los trabajadores están cambiando, se buscan nuevas oportunidades que satisfagan mejor sus expectativas de salario, con mayor proyección de crecimiento y que tengan mejor equilibrio entre vida laboral y personal (Cuesta,2023).

Asimismo, un estudio realizado en México trata de descifrar las principales causas del aumento de los índices de rotación de personal entre los jóvenes. Las causas predominantes identificadas fueron la falta de oportunidades de crecimiento, el descontento con el liderazgo y el ambiente laboral poco estimulante (Cárdenas,2023).

Ambos estudios sustentan la hipótesis generada a través de nuestra investigación que soporta que los jóvenes tienen nuevas inquietudes y preferencias a la hora de buscar un puesto de trabajo, lo que da lugar a un aumento en los niveles de renuncia voluntaria.

A continuación, se buscan informes que sustenten la relación establecida por nuestro análisis de datos entre la cantidad de horas extra y la renuncia voluntaria. Según la investigación a mayor número de horas extras realizadas por el empleado más aumenta la probabilidad de que el empleado renuncie a su puesto de trabajo de manera voluntaria. Una investigación publicada por “Pew Research Center” pone de manifiesto que aproximadamente el 39% de los trabajadores que abandonaron su puesto de trabajo en 2021 establecieron el trabajar demasiadas horas como una de las principales razones por las que decidieron abandonar (Blazina, 2022).

Adicionalmente, un estudio llevado publicado por JSTOR señala que las horas extra están asociadas con un aumento de los niveles de rotación. Principalmente por las consecuencias que esto puede acarrear con relación a problemas de salud tanto físico como mentales, o disminución de satisfacción laboral (Bannai & Tamakoshi, 2019).

Por último, un estudio realizado en colaboración de la Organización Mundial de la Salud (OMS) y la Organización del Trabajo (OIT). Puso de manifiesto la relación entre la carga de trabajo excesiva y cómo esta contribuye a la decisión por parte de los trabajadores de dejar voluntariamente sus empleos. A fin de conseguir un mayor equilibrio entre vida laboral y personal. También se destaca en el efecto negativo en la

salud y bienestar de los trabajadores como uno de los principales móviles para tomar la decisión de renuncia voluntaria (Johnson,2021).

Todas estas investigaciones dan validez a las conclusiones que hemos obtenido en nuestro análisis.

Seguidamente, se identifican posibles investigaciones que avalen la hipótesis obtenida en relación con la satisfacción laboral y la renuncia voluntaria. El estudio pone de manifiesto que cuanto mayor sea el grado de satisfacción laboral menor será el grado de rotación por parte de los empleados de la empresa. Una investigación llevada a cabo por la empresa “McKinsey & Company” expone como muchos empleados dejaron sus trabajos motivados por los ambientes tóxicos de trabajo. El verse expuestos a una continua sensación de infravaloración e insatisfacción empujó a los trabajadores a buscar otras empresas con mejores condiciones (De Smet et al. ,2022).

Asimismo, uno de los informes que se mencionó ya anteriormente, realizado por “Pew Research Center “señala que la principal razón que motiva el fenómeno de renuncia voluntaria es la insatisfacción con las condiciones laborales y ambiente de trabajo. El 57% de los trabajadores señaló el sentimiento de falta de respeto como factor crítico a la hora de tomar la decisión de abandonar su puesto de trabajo (Blazina,2022).

Además, un estudio publicado en “U.S Bureau of labor statistics (BLS)” reporta que en aquellos sectores donde los trabajadores se sienten menos valorados y el grado de satisfacción con el ambiente de trabajo es menor, las tasas de renuncia son más altas (Gittleman,2022).

Por tanto, resulta evidente que tal y como pone de manifiesto nuestra investigación los bajos niveles de satisfacción laboral generan un aumento de la rotación en las organizaciones.

Por último, se trata de verificar las conclusiones obtenidas en nuestro estudio en referencia al puesto de trabajo y la renuncia voluntaria. Primeramente, se han encontrado investigaciones que dan soporte a una de las principales conclusiones, que el 40% de los representantes de ventas renuncia de manera voluntaria.

Según la investigación publicada por “The European Data Journalism Network (EDJNet)”, en varios países europeos, las renunciaciones voluntarias han aumentado

considerablemente. Específicamente, en el sector de las ventas. A causa de ciertas condiciones laborales y el estrés asociado al puesto (Finizio,2023).

A su vez, otro estudio publicado de la mano de “Emerald Insight” pone de manifiesto que los trabajadores en el sector de ventas son particularmente propensos a renunciar voluntariamente en caso de insatisfacción laboral por condiciones laborales o falta de expectativas de crecimiento. Adicionalmente, vuelven a comentar el estrés y la presión como factores que contribuyen a la alta rotación (Ng & Stanton, 2023).

Ambos estudios apoyan la conclusión obtenida en este trabajo de investigación entre la relación de los trabajadores pertenecientes al sector ventas y el factor de renuncia voluntaria.

Tras esta exhaustiva investigación por diversos estudios publicados que giran en torno a la renuncia voluntaria en las empresas y sus principales causas se ha demostrado la veracidad de los “insights” obtenidos tras realizar el trabajo de investigación.

Es fundamental contrastar las conclusiones obtenidas para asegurar que el análisis subsiguiente sobre las posibles implicaciones prácticas del modelo se base en fundamentos sólidos y bien verificados. Este proceso de verificación no solo refuerza la validez de nuestras conclusiones, sino que también proporciona una base confiable para la toma de decisiones informadas. Además, una revisión exhaustiva de las conclusiones permite identificar posibles sesgos o errores, garantizando así que cualquier recomendación práctica derivada del análisis sea robusta y aplicable en contextos reales.

5.2 Implicaciones prácticas del modelo

Una vez concluido el trabajo de investigación, es crucial identificar las posibles implicaciones prácticas que puede tener el modelo al implementarse en empresas reales. Contar con un modelo que permita, por un lado, identificar las variables que más influyen en la decisión de los empleados de renunciar voluntariamente y, por otro lado, detectar a aquellos trabajadores con mayor probabilidad de hacerlo, abre un abanico de posibilidades significativas. Este conocimiento puede transformar de manera extraordinaria la gestión de recursos humanos en las empresas. Entre las principales implicaciones prácticas se incluyen:

- Desarrollo de estrategias de retención personalizadas: Al ser capaces de identificar a los empleados más propensos a tomar la decisión de renunciar, las empresas pueden diseñar estrategias específicas para abordar estas preocupaciones. Ofreciendo incentivos personalizados o soluciones que aborden sus inquietudes antes de que los trabajadores identificados decidan abandonar la organización. Por ejemplo, imaginemos que un empleado que aspira a tener un rol de liderazgo es detectado por nuestro modelo como individuo que se está planteando renunciar a su puesto. En tal caso, la empresa podría implementar una estrategia de retención personalizada ofreciéndole mentorías con altos directivos, programas de formación en liderazgo y oportunidades de asumir responsabilidades adicionales. Tratando así de hacer cambiar de opinión al empleado (Castro et al., 2022).
- Aumento y mejora de planes de carrera: Según se ha identificado en el modelo una gran mayoría de personas que renuncian a su puesto en la empresa son jóvenes. Principalmente porque no ven oportunidades de crecimiento y desarrollo de sus habilidades en esa organización. Este valioso "insight" sugiere que los departamentos de recursos humanos deben reevaluar y mejorar sus planes de carrera para los empleados.

Para abordar esta problemática, es crucial diseñar e implementar planes de carrera más personalizados, que se adapten a las inquietudes y aspiraciones individuales de cada empleado (Monsalve,2020). Estos planes deben incluir oportunidades claras y accesibles para el desarrollo profesional y la adquisición

de nuevas competencias. Además, es fundamental complementar estos planes con programas de formación continua que permitan a los empleados mejorar y actualizar sus habilidades en función de las demandas del mercado y los objetivos estratégicos de la empresa.

La personalización de los planes de carrera debe basarse en un entendimiento profundo de las metas profesionales de los empleados, obtenido a través de evaluaciones periódicas y conversaciones abiertas sobre sus aspiraciones. Al proporcionar una hoja de ruta clara para el avance profesional, las empresas pueden aumentar significativamente la satisfacción y el compromiso de sus empleados jóvenes, reduciendo así la tasa de rotación.

- **Mejora clima laboral:** El modelo predictivo ha identificado que uno de los factores más influyentes en la decisión de los empleados de renunciar voluntariamente es la satisfacción laboral. Diversos elementos contribuyen a esta satisfacción, y es fundamental que las empresas los aborden de manera integral para mejorar el clima laboral y retener talento. En primer lugar, el trato que los empleados reciben de sus superiores es crucial. Un liderazgo que demuestre respeto, apoyo y reconocimiento puede significativamente aumentar la satisfacción laboral. Los empleados necesitan sentir que su trabajo es valorado y que sus contribuciones son reconocidas regularmente. Además, la gestión del estrés laboral es esencial. Las empresas deben evaluar y, en la medida de lo posible, reducir las fuentes de estrés, como cargas de trabajo excesivas y plazos irrazonables. Proporcionar recursos para la gestión del estrés, como programas de bienestar y apoyo psicológico, puede ser muy beneficioso. Adicionalmente, las horas extra no remuneradas o excesivas también afectan negativamente la satisfacción. Las políticas de equilibrio entre la vida laboral y personal deben ser claras y justas, promoviendo un entorno donde los empleados no se sientan obligados a sacrificar su tiempo personal. Asimismo, las instalaciones de la oficina juegan un papel importante en el bienestar diario de los empleados. Espacios de trabajo cómodos, bien diseñados y equipados adecuadamente pueden mejorar significativamente la moral y la productividad. Además, la existencia de beneficios y prestaciones para los empleados, como programas de bienestar, vales de restaurante, membresías de

gimnasio y otros, también contribuyen positivamente a la satisfacción laboral. Finalmente, fomentar un entorno de trabajo colaborativo y de apoyo, donde se valore y promueva el trabajo en equipo, es vital. La creación de una cultura corporativa inclusiva y cooperativa puede fortalecer las relaciones entre los empleados y mejorar el clima general de la organización (Haider et al. ,2019).

- Mejora de sistemas de recompensas: Por último, el modelo predictivo ha identificado que los salarios y beneficios son factores clave para la retención de empleados. Por lo tanto, es esencial que las organizaciones revisen y optimicen sus sistemas de recompensas para asegurar que el esfuerzo de su capital humano esté adecuadamente compensado, alineado con el sector (García et al., 2012). La compensación directa debe ser competitiva y reflejar el valor del trabajo realizado. Esto incluye ofrecer sueldos que sean competitivos dentro del sector, basados en análisis de mercado y benchmarking; implementar programas de bonos que recompensen el desempeño individual y grupal, incentivando el cumplimiento de objetivos y la excelencia; y establecer estructuras de comisiones atractivas que motiven a los empleados a alcanzar y superar sus metas.

La compensación indirecta también juega un papel crucial en la satisfacción y retención de empleados. Esto puede incluir ofrecer planes de pensiones o contribuciones a fondos de jubilación, proporcionando seguridad a largo plazo para los empleados; proveer seguros de salud completos que cubran necesidades médicas, dentales y de visión, mostrando un compromiso con el bienestar de los empleados; establecer políticas generosas de vacaciones pagadas que permitan a los empleados desconectar y recargar energías; e incluir beneficios adicionales como membresías de gimnasio, vales de restaurante, programas de bienestar mental, y opciones de trabajo flexible, que contribuyan a un equilibrio saludable entre la vida laboral y personal.

Para que un sistema de recompensas sea efectivo, es fundamental que sea justo y transparente. Los empleados deben entender claramente cómo se calculan y distribuyen las recompensas, y percibir que el sistema es equitativo. Asimismo, el sistema debe ser adaptable y dinámico, revisado y ajustado periódicamente para

adaptarse a los cambios en el mercado laboral, las condiciones económicas y las necesidades de los empleados.

Esta serie de recomendaciones prácticas dirigidas a las empresas tras analizar los resultados del modelo, pretenden mejorar la tasa de retención de empleados consiguiendo así la disminución de los costes asociados a altos niveles de rotación. A modo resumen, entre las posibles implicaciones prácticas destacan el desarrollo de estrategias de retención personalizadas, la mejora de las proyecciones y planes de carrera ofrecidos en la empresa, el progreso del clima laboral y la revisión del sistema de recompensas.

6 CONCLUSIONES

Tras finalizar el trabajo de investigación se establecen una serie de conclusiones en forma de cierre del presente trabajo.

El trabajo busca integrar Business Analytics en las empresas para resolver problemas y optimizar procesos, destacando la importancia del análisis de datos y la inteligencia generativa para mantener la competitividad. Se centra en recursos humanos, utilizando "People Analytics" para mejorar la gestión del talento y la retención de personal. Motivado por la alta rotación de personal, el estudio identifica los costos asociados y propone un modelo predictivo de renuncia voluntaria. El objetivo es desarrollar estrategias que mejoren la retención y satisfacción de los empleados.

El área de Recursos Humanos (RRHH) gestiona a los empleados, evolucionando de tareas administrativas a responsabilidades estratégicas. La gestión de RRHH incluye selección, formación, evaluación, promoción, y motivación del personal. Los objetivos varían según la empresa, siendo más complejos en grandes corporaciones, y se centran en la eficiencia económica, social y coherencia de acciones. La digitalización de RRHH en España usa Big Data, computación en la nube, redes sociales, movilidad y gamificación para mejorar la gestión del talento en selección, formación, administración y compensación, aumentando la eficiencia y satisfacción laboral. Asimismo, el "Machine Learning" en RRHH analiza grandes datos para optimizar la selección y retención de empleados, prediciendo renuncias y permitiendo intervenciones proactivas para reducir costos.

A continuación, se construyó un modelo explicativo para predecir si un empleado renuncia voluntariamente usando regresión logística. Este análisis permite identificar variables clave que influyen en esta decisión, como la satisfacción laboral, el balance entre trabajo y vida personal, y la cantidad de horas extras trabajadas. Se realizó un primer modelo con todas las variables y luego se ajustó usando aquellas con mayor influencia. Posteriormente, se construyó un primer modelo predictivo con regresión logística, aunque presentó problemas con falsos negativos. Se intentó mejorar el modelo usando "stepwise" para seleccionar variables.

Finalmente, se utilizó el algoritmo de Random Forest, que demostró mejor rendimiento, mejorando significativamente la sensibilidad. Comparativamente, Random Forest fue más efectivo que la regresión logística en predecir renuncias voluntarias, identificando correctamente el 80% de los casos.

Seguidamente, en el análisis de resultados se destacan tres modelos predictivos. No obstante, el modelo "Random Forest" mostró un mejor desempeño, por lo que nos enfocamos en este para identificar las variables más influyentes en la decisión de un empleado de abandonar su trabajo. Se utilizaron diversos gráficos para visualizar la relación entre estas variables y la decisión de renunciar voluntariamente, revelando patrones significativos y ayudando a identificar factores clave que influyen en la retención de empleados.

Por último, La discusión valida los hallazgos comparando el análisis de datos con la literatura existente. Adicionalmente, a efectos de contribución social y organizativas se resaltan especialmente las implicaciones prácticas que en el capítulo 5.2. Se incluyen estrategias de retención personalizadas, mejorar planes de carrera y el clima laboral, reducir el estrés, y optimizar los sistemas de recompensas. Estas estrategias pueden mejorar la gestión de recursos humanos, aumentar la retención y reducir costos asociados a la alta rotación.

7 DECLARACIÓN RESPECTO AL USO DE CHAT GPT U OTRAS HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA.


ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Cristina Hernández Ayúcar, estudiante de Administración y dirección de empresas y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Proyecto de emprendimiento Uptoyou", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación.

1. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
2. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
3. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 10/06/2024

Firma:  Cristina Hernández Ayúca

8 BIBLIOGRAFÍA

Bannai, A., & Tamakoshi, A. (2019). The association between long working hours and health: A systematic review of epidemiological evidence on JSTOR. www.jstor.org, 40(1).

<https://www.jstor.org/stable/43187983#:~:text=URL%3A%20https%3A%2F%2Fwww.jstor.org%2Fstable%2F43187983%0A%5BImage%20%5D%0AVisible%3A%20%25%20,100>

Blazina, C. (2022, marzo 9). Majority of workers who quit a job in 2021 cite low pay, no opportunities for advancement, feeling disrespected. Pew Research Center. <https://www.pewresearch.org/short-reads/2022/03/09/majority-of-workers-who-quit-a-job-in-2021-cite-low-pay-no-opportunities-for-advancement-feeling-disrespected/>

Bonillo Muñoz, D. (2002). La satisfacción laboral como elemento motivador del empleado. *Trabajo*, 11. <https://doi.org/10.33776/trabajo.v11i0.166>

Cárdenas, L. (2023, junio 27). Jóvenes renuncian a sus trabajos en México por estas 4 razones. MVS Noticias. <https://mvsnoticias.com/entrevistas/2023/6/27/jovenes-renuncian-sus-trabajos-en-mexico-por-estas-razones-597334.html>

Castillo Pizo, E. (2021). El impacto de la medida de rotación de los empleados en las organizaciones. <https://repository.unicatolica.edu.co/handle/20.500.12237/2208>

Castro, A. C., et al. (2022). Implementación de machine learning en la rotación de personal [Trabajo de grado, Universidad EAN]. Recuperado de: <http://hdl.handle.net/10882/13431>.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>

Cuesta, L. (2023, marzo 29). La generación Z desvinculada de su trabajo: ¿Por qué los jóvenes quieren dejar sus empleos? *La Vanguardia*. <https://www.lavanguardia.com/vida/formacion/20230329/8855916/generacion-z-desvinculada-trabajo-jovenes-quieren-dejar-empleos-mkt-emg.html>

Cutler, A., Liaw, R. P. by, & Wiener., M. (2022, mayo 23). VarImpPlot: Variable importance plot in randomForest: Breiman and cutler's random forests for classification and regression. Rdr.io. <https://rdr.io/cran/randomForest/man/varImpPlot.html>

DataScientest. (2022, enero 25). Random Forest: Bosque aleatorio. Definición y funcionamiento. Formación en ciencia de datos | DataScientest.com; DataScientest. <https://datascientest.com/es/random-forest-bosque-aleatorio-definicion-y-funcionamiento>

De Smet, A., Dowling, B., Hancock, B., & Schaninger, B. (2022, julio 13). The Great Attrition is making hiring harder. Are you searching the right talent pools? Mckinsey.com; McKinsey & Company. <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-great-attrition-is-making-hiring-harder-are-you-searching-the-right-talent-pools>

EAE Business School. (2023, octubre 27). ¿Qué es el People Analytics y cómo se integra la herramienta de datos en RRHH? EAE Business School. <https://www.eaebarcelona.com/es/blog/people-analytics-herramienta>

Estrada, P. R. El costo de la rotación de personal. Deloitte. Recuperado el 28 de mayo de 2024, de https://f.hubspotusercontent00.net/hubfs/9188006/Documentos%20Blog/Costo%20de%20Rotacio%CC%81n%20Personal_2018.pdf

Estudio employer brand research. (2023, septiembre 27). Randstad. <https://www.randstad.es/estudio-employer-brand-research/>

Fernández Palma, M. D. (2009). La evaluación de desempeño, la percepción de justicia y las reacciones de los empleados. *Contabilidad y negocios*, 4(8), 42–48. <https://doi.org/10.18800/contabilidad.200902.005>

Fernández Rodríguez, L., Perona, M., Martínez, F., Reichardt, S., Del Rio, J., & Martínez, L. (2018). La digitalización de la función de RR.HH. en España. *Capital Humano*, 327.

Fernández Rodríguez, L., Perona, M., Martínez, F., Reichardt, S., Del Rio, J., & Martínez, L. (2021). Observatorio sobre la digitalización de la función de RR.HH. en España. *Capital Humano*.

Finizio, M. (2023, marzo 28). The great turnover: record resignations and job vacancies in Europe. European Data Journalism Network - EDJNet. https://www.europeandatajournalism.eu/cp_data_news/the-great-turnover-record-resignations-and-job-vacancies-in-europe/

Fitz-Enz, J., & Mattox, J., II. (2014). Predictive Analytics for Human Resources. John Wiley & Sons.

Gallastegi, E. A., & Rodríguez, J. L. (2013). Dirección Estratégica de los Recursos Humanos: teoría y práctica. Ediciones Pirámide, 21-28. <https://dialnet.unirioja.es/servlet/libro?codigo=463551>

García Pérez, R., García Pino, G., González Ballester, D., & García Moreno, R. (2010). Modelo de regresión logística para estimar la dependencia según la escala de Lawton y Brody. *Semergen*, 36(7), 365–371. <https://doi.org/10.1016/j.semereg.2010.03.004>

García.M, Iveth.N, Hernández.C.F. (2012). La motivación y los sistemas de recompensas y su impacto en la producción. *Dialnet*, 10(2). <https://dialnet.unirioja.es/servlet/articulo?codigo=9173769>

Gittleman, M. (2022). The “Great Resignation” in perspective. *Bls.gov*. <https://www.bls.gov/opub/mlr/2022/article/the-great-resignation-in-perspective.htm>

Grisales, R., & Viviana, P. (2020). El impacto que causa la alta rotación de personal en las organizaciones. <https://repository.unimilitar.edu.co/handle/10654/36209>

Guerrero Velástegui, C., & Parra Real, S. (2023). Clima laboral y su relación con la rotación del personal [Universidad Técnica de Ambato]. <https://repositorio.uta.edu.ec/jspui/handle/123456789/37202>

Haider, S., Ahmed, M., Anwer, S., Aslam, M., & Ali Asadullah, M. (2019). Efecto de la coordinación relacional en las intenciones de rotación de empleados a través de la satisfacción laboral: el uso de ecuaciones estructurales y la simulación de Monte Carlo. *ESIC MARKET Economic and Business Journal*, 50(162), 43–65. <https://doi.org/10.7200/esicm.162.0501.1>

HR-Employee-Attrition.csv. (2017). [Data set]. En IBM HR Analytics Employee Attrition & Performance. <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/data>

Huanuco Naupay, E. J., & Santillan Carrion, J. M. (2023). La marca del empleador y rotación de personal en una empresa de transporte refrigerado. Universidad César Vallejo.

IBM. (2021, diciembre 7). Regresión logística binaria SPSS statistics. Ibm.com. <https://www.ibm.com/docs/es/spss-statistics/beta?topic=regression-binary-logistic>

IBM. (2024, mayo 10). ¿Qué es la regresión logística? Ibm.com. <https://www.ibm.com/es-es/topics/logistic-regression>

ISDI. (2023, julio 8). HR Analytics: qué es y cómo aplicarlo en tu empresa. ISDI España; ISDI España | La Escuela de Negocios de la Era Digital. <https://www.isdi.education/es/blog/hr-analytics-aplicacion-en-empresas>

Johnson, C. (2021). La OMS y la OIT alertan de que las jornadas de trabajo prolongadas aumentan las defunciones por cardiopatía isquémica o por accidentes cerebrovasculares. Who.int. <https://www.who.int/es/news/item/17-05-2021-long-working-hours-increasing-deaths-from-heart-disease-and-stroke-who-ilo>

Lipton, Z. C., Elkan, C., & Narayanaswamy, B. (2014). Thresholding classifiers to maximize F1 score. En arXiv [stat.ML]. <http://arxiv.org/abs/1402.1892>

López Cháñez, F. J., Casique Guerrero, A., & Ferrer Guerra, J. (2007). La satisfacción hacia el trabajo. Un análisis basado en las teorías de las expectativas y de equidad. Entelequia. Revista Interdisciplinar, 3, 219–234. <https://ideas.repec.org/a/erv/ancoec/y2007i3p219-234.html?ref=Sex%C5%9Ehop.Com>

López, J (2011). Propuesta para reducir la rotación de personal. Revista Vinculando, 9(1). https://vinculando.org/empresas/propuesta_para_reducir_el_indice_de_rotacion_de_personal.html

Martín, D. (2023, abril 25). People Analytics: Las últimas tendencias de Big Data y Business Intelligence aplicadas a los RRHH. - BDO. Bdo.es. <https://www.bdo.es/es-es/blogs-es/coordenadas-bdo/las-ultimas-tendencias-aplicadas-a-los-rrhh>

Martín Sierra, C. (2011). Gestión de recursos humanos y retención de capital humano estratégico: análisis de su impacto en los resultados de empresas innovadoras españolas. Dialnet. <https://dialnet.unirioja.es/servlet/tesis?codigo=296085>

Martínez, Diego & Albín, Julio & Cabaleiro, José & Pena, Tomas & Rivera, Francisco & Blanco, Vicente. (2009). El criterio de información de Akaike en la obtención de modelos estadísticos de Rendimiento. 439-444.

Marrugo, M y Pérez Meléndez, B. (2012). Análisis de la teoría de las expectativas de Victor Vroom en el centro de emprendimiento y desarrollo Pedro Romero "Cemprende". Universidad de Cartagena.

Monsalve, M. L. (2020). MODELO PREDICTIVO PARA IDENTIFICAR FACTORES QUE INFLUYEN EN LA ROTACIÓN DEL PERSONAL EN UNA ORGANIZACIÓN DEL SECTOR [Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/bitstream/handle/unal/77442/1035832177.2019.pdf?sequence=4&isAllowed=y>

Navarro, Guillermo. (2022). Analítica de RRHH: Machine Learning.

Ng, E., & Stanton, P. (2023). Editorial: The great resignation: managing people in a post COVID-19 pandemic world. *Personnel Review*, 52(2), 401–407. <https://doi.org/10.1108/pr-03-2023-914>

One hot encoding. (2023). Interactivechaos.com. <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/one-hot-encoding>

Peeters, T., Paauwe, J., & Van De Voorde, K. (2020). People analytics effectiveness: developing a framework. *Journal of Organizational Effectiveness People and Performance*, 7(2), 203–219. <https://doi.org/10.1108/joepp-04-2020-0071>

Perez, M. (2023, julio 13). People Analytics: qué es y aplicaciones en Recursos Humanos. SMOWL Proctoring | Sistema de supervisión para exámenes online. <https://smowl.net/es/blog/people-analytics/>

Puyol-Cortez, J. L. (2021). La satisfacción laboral como factor clave para la productividad en las PYMES de Santo Domingo de los Tsáchilas. *Journal of Economic and Social Science Research*, 1(2), 50–64. <https://doi.org/10.55813/gaea/jessr/v1/n2/31>

Qué es el análisis predictivo de RRHH: Guía para RRHH y selección de personal. (s. f.). <https://www.manatal.com/es/blog/predictive-hr-analytics>

Quintero Mejía, J. J., & Ordoñez Barajas, M. (2023). Análisis del comportamiento de los factores asociados a la intención de la rotación en empleados de empresas manufactureras. <http://repositorio.uts.edu.co:8080/xmlui/handle/123456789/12337>

Randstad. (2016). El salario, el factor más importante a la hora de elegir dónde trabajar. Fororecursoshumanos.com. <https://www.fororecursoshumanos.com/salario-factor-mas-importante-hora-elegir-donde-trabajar/>

RETENCIÓN DE PERSONAL - ingenia. (2020, 7 enero). Ingenia. <https://ingenia.com/retenciondepersonal/>

Roy-García, I. A., Paredes-Manjarrez, C., Moreno-Palacios, J., Rivas-Ruiz, R., & Flores-Pulido, A. A. (2023). ROC curves: general characteristics and their usefulness in clinical practice. En *Revista Médica del Instituto Mexicano del Seguro Social* (Vol. 61, Número Suppl 3, p. S497). Zenodo. <https://doi.org/10.5281/ZENODO.8319791>

Salazar, T. (2022). IDENTIFICACIÓN DE LOS FACTORES QUE DETERMINAN LA ALTA ROTACIÓN DEL PERSONAL DE VENTAS EN EMPRESAS RETAIL DE PRENDAS DE VESTIR: CASO TIENDAS PIERO [Universidad de Lima]. https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/1724/Salazar_Prado_Talia.pdf?sequence=1&isAllowed=y

Sanahuja, P. M., & de Pol Martí Sanahuja, V. T. las E. (2021). Entendiendo la curva ROC y el AUC: Dos medidas del rendimiento de un clasificador binario que van de la mano. Polmartisanahuja.com. Recuperado de <https://polmartisanahuja.com/entendiendo-la-curva-roc-y-el-auc-dos-medidas-del-rendimiento-de-un-clasificador-binario-que-van-de-la-mano/>

Torres, D. (2023, junio 21). Rotación de personal: qué es, causas, consecuencias y soluciones. Hubspot.es. <https://blog.hubspot.es/sales/rotacion-personal>

Tovar, J. C. (2022, septiembre 10). Aprendizaje Automático - One Hot Encoding para datos categóricos. Huawei.com. <https://forum.huawei.com/enterprise/es/Aprendizaje-Autom%C3%A1tico-One-Hot-Encoding-para-datos-categ%C3%B3ricos/thread/667232037213454336-667212895009779712>

Tursunbayeva, A., Di Lauro, S., & Pagliari, C. (2018). People analytics—A scoping review of conceptual boundaries and value propositions. *International Journal of Information Management*, 43, 224–247. <https://doi.org/10.1016/j.ijinfomgt.2018.08.002>

Variable importance plot using random forest package in R. (2022, noviembre 19). GeeksforGeeks. <https://www.geeksforgeeks.org/variable-importance-plot-using-random-forest-package-in-r/>

Villalba.F. (2023). Capítulo 6 Regresión logística binaria. Github.io.
<https://fervilber.github.io/Aprendizaje-supervisado-en-R/glm.html>

Wolterskluwer. (2023, marzo 13). People analytics qué es y cómo gestionar los datos en RRHH. Wolterskluwer.com. <https://www.wolterskluwer.com/es-es/expert-insights/people-analytics-datos-rrhh>

