



TRABAJO DE GRADO
Opción Seminario-Diplomado.

Propuesta de trabajo de grado machine learning

**Algoritmo computacional para la predicción de retiros en un entorno
hospitalario**

Corporación Universitaria Remington.
Facultad de ingenierías
Seminario ruta crehana ingeniería

Estudiantes:
Santiago Agudelo Jiménez
Santiago Castañeda Arango

Tutor:
Juan Pablo Velez Uribe

Opción de Trabajo de grado Seminario-Diplomado.
2023

Agradecimientos

Juan Pablo Velez Uribe
Analytics DATA
Clínica SOMER

Tabla de Contenidos

Resumen.....	4
Palabras clave.....	5
Marco conceptual y textual.....	5
Conceptos de Introducción	5
Machine Learning	5
Big Data	6
Redes Sociales	6
Comercio Electrónico	6
Salud y Ciencias Biológicas.....	6
Finanzas	6
Telecomunicaciones.....	7
Industria Manufacturera.....	7
Transporte y Logística	7
Gobierno	7
Roles y habilidades	7
Pirámide de Valor	8
Librerías en Python.....	8
Aprendizaje supervisado y no supervisado.....	9
Análisis contrafactual.....	10
Contexto del Trabajo.....	11
Descripción de la Organización	12
Ética en Machine Learning	13
Desarrollo e implementación del aprendizaje.....	14
Hablemos de XGBoost	15
¿Por qué QlikSense?	16
BI en la inteligencia artificial.....	17
Métricas.....	17
Apendice	18
Conclusiones	27
Bibliografía	28

Referencias.....	3
	29

Resumen

El presente trabajo de grado se enfoca en afrontar el desafío significativo que representa la elevada tasa de retiros de personal en un entorno hospitalario. Esta problemática se refleja en la pérdida considerable de tiempo y recursos derivados del constante ciclo de búsqueda, selección y formación de nuevos empleados. Además, repercute directamente en la calidad de la atención médica ofrecida a los pacientes. El objetivo principal abarca no solo comprender las razones subyacentes que impulsan estos retiros laborales, sino también proponer y desarrollar estrategias concretas y eficaces para retener al personal existente. El propósito último es mitigar las salidas laborales mensuales, reduciendo así el impacto negativo en las operaciones diarias de un hospital.

El estudio se dirige a investigar las causas profundas de los retiros del personal hospitalario, considerando factores diversos como el ambiente laboral, las condiciones laborales, las oportunidades de crecimiento profesional y otros elementos relevantes que puedan influir en la decisión de los empleados de abandonar sus puestos. Este análisis exhaustivo busca identificar patrones y tendencias que puedan estar contribuyendo a la alta rotación del personal.

Además de la identificación de las causas, se plantea el diseño de estrategias de retención del personal basadas en un enfoque proactivo y preventivo. Estas estrategias tienen como objetivo no solo retener a los empleados actuales, sino también promover un ambiente laboral más satisfactorio y estimulante que fomente la permanencia del talento dentro del hospital. Se espera que la implementación de estas estrategias ayude a minimizar la constante rotación de empleados y, en consecuencia, reduzca la interrupción en las operaciones hospitalarias, mejorando así la continuidad y calidad en la atención brindada a los pacientes.

Palabras clave

Retención de personal, Hospital, Rotación laboral, Estrategias de retención, Recursos humanos, Desarrollo, Machine learning, Datos

Marco conceptual y textual

Conceptos de Introducción

El presente informe técnico se centra en la problemática de la retención de personal en un entorno hospitalario, abordando conceptos fundamentales relacionados con la gestión de recursos humanos en el sector de la salud. Además, se explora como la aplicación de tecnologías como el Machine Learning puede potenciar las estrategias de retención y mejorar la eficiencia en la gestión del personal. Se consideran conceptos clave como la retención de empleados, la rotación laboral, estrategias de retención y la importancia de un entorno laboral satisfactorio.

Machine Learning

El Machine Learning o aprendizaje automático, es una rama de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las máquinas aprender de datos y realizar tareas sin ser programadas explícitamente. En lugar de seguir instrucciones específicas, los sistemas de Machine Learning utilizan patrones y experiencias previas para mejorar su rendimiento a medida que se exponen a más información. [1]

En la actualidad muchas aplicaciones que utilizamos frecuentemente aplican el Machine Learning sin nosotros tener el conocimiento Ejemplo:

- Recomendaciones personalizadas, las grandes plataformas de comercio electrónico y streaming utilizan algoritmos para analizar el comportamiento del usuario y ofrecer recomendaciones personalizadas.
- Búsquedas avanzadas, motores de búsqueda como Google emplean técnicas de machine Learning para perfeccionar los resultados de búsqueda.
- Detección de fraudes, plataformas financieras y de comercio aplican algoritmos de machine Learning para analizar patrones de transacciones y comportamiento de usuario identificando posibles actividades fraudulentas

- Optimización publicitaria, Las plataformas publicitarias emplean algoritmos de machine Learning para personalizar anuncios según el comportamiento y las preferencias del usuario.

Big Data

Este concepto se refiere fundamentalmente al uso masivo de datos que se almacenan y utilizan en las bases de datos y que se utilizan. La aparición de nuevas tecnologías, como la computación en la nube, el denominado internet de las cosas permiten que se vaya generando una creciente cantidad de datos relativos al comportamiento de los individuos. Estos conjuntos de datos no solo son grandes en tamaño, sino que también presentan desafíos en términos de variedad, velocidad y veracidad. La gestión y análisis efectivos de Big Data requieren tecnologías avanzadas que permitan almacenar, procesar y extraer información valiosa de estos enormes conjuntos de datos.

Muchas grandes empresas manejan grandes volúmenes de datos a continuación algunas de ellas.

Redes Sociales

Plataformas como Facebook y Twitter manejan grandes volúmenes de datos en tiempo real, incluyendo publicaciones, comentarios, y actividades de usuarios. El análisis de estos datos ayuda a personalizar las experiencias de los usuarios y a mejorar la orientación publicitaria.

Comercio Electrónico

Empresas como Amazon utilizan Big Data para analizar patrones de compra, comportamientos de navegación y preferencias del cliente. Esto facilita la recomendación de productos personalizados y la optimización de la cadena de suministro.

Salud y Ciencias Biológicas

En la investigación médica, el Big Data se utiliza para analizar grandes conjuntos de datos genómicos, datos clínicos y registros de pacientes. Esto ayuda en la identificación de patrones, descubrimiento de medicamentos y personalización de tratamientos.

Finanzas

Instituciones financieras gestionan grandes cantidades de datos relacionados con transacciones, riesgos y comportamientos del mercado. El análisis de Big Data en finanzas se aplica para la detección de fraudes, gestión de riesgos y toma de decisiones basada en datos.

Telecomunicaciones

Proveedores de servicios de telecomunicaciones analizan grandes volúmenes de datos de llamadas, mensajes y uso de datos para mejorar la calidad del servicio, predecir congestiones de red y personalizar ofertas para los usuarios.

Industria Manufacturera

En entornos de fabricación, el Big Data se utiliza para monitorear el rendimiento de las máquinas, prevenir fallas, optimizar la cadena de suministro y mejorar la eficiencia operativa.

Transporte y Logística

Compañías de transporte y logística emplean Big Data para rastrear vehículos, optimizar rutas de entrega, prever mantenimientos y mejorar la eficiencia en la gestión de flotas.

Gobierno

Las agencias gubernamentales utilizan Big Data para analizar datos demográficos, mejorar la toma de decisiones políticas, prever patrones delictivos, y optimizar la prestación de servicios públicos.

Roles y habilidades

Un equipo de ciencia de datos exitoso debe contar con una combinación equilibrada de roles y habilidades para abordar los diversos aspectos del ciclo de vida de los datos, si se desea realizar un proyecto de machine Learning se debe contar con las personas aptas para realizar las tareas propuestas, muchas empresas piensan que con solo el científico de datos pueden sacar a flote un proyecto, pero la realidad es otra, existen un equipo completo de personas que se encarga de dar valor al proyecto.

- **Científico de Datos:** Un científico de datos es un profesional dedicado a analizar e interpretar grandes bases de datos, cuenta con habilidades en programación, estadística, machine Learning [3]
- **Ingeniero de Datos:** Un ingeniero de datos se encarga de diseñar y construir sistemas de recogida, almacenamiento y análisis de datos a gran escala, cuenta con habilidades en SQL, ETL, Automatización [4]
- **Machine Learning Engineer:** Un ingeniero en machine Learning es parecido a un científico de datos, pero esta persona posee más experiencia en implementación y despliegue de modelos de machine Learning posee conocimientos en frameworks como tensorflow o pytorch, habilidades en optimización y ajuste de modelos
- **Personas de Negocios:** Estas personas son muy importante ya que con ellas nos ayudan a enfocar el proyecto por el camino correcto, estas personas conocen como funciona la empresa.

Pirámide de Valor

La pirámide de valor de los datos es una analogía que se utiliza para estructurar los pasos de una estrategia a seguir en la importancia de los datos. [5]

- **Tecnología:** Esta es la base de la pirámide ya que sin la tecnología correcta no podremos procesar los datos, una empresa deberá realizar una buena inversión en tecnología.
- **Gobernanza de los Datos:** Las tecnologías y datos crudos necesitan equipo y prácticas de seguridad, calidad, documentación y manejo de la información.
- **Operaciones Fundamentales:** en esta etapa la información permite realizar transacciones como compras en una pagina web o reportes básicos financieros.
- **Inteligencia de Negocios:** Herramientas que permitan análisis descriptivos y visualizaciones con dashboard o tableros de control, permiten responder preguntas de una manera más profunda.
- **Inteligencia Artificial:** Procesos de decisión automatizados con inteligencia artificial, generando valor a través de los algoritmos.

Un mensaje claro de esta pirámide es para poder llegar a la punta no es posible tomar atajos debemos construirla desde la base. El machine Learning es una tecnología que depende de los datos y sin una buena inversión en tecnología y procesos no podremos realizar machine Learning.

Librerías en Python

En el ámbito de la inteligencia artificial en Python, hay varias bibliotecas importantes que se utilizan ampliamente para diversas tareas, desde el aprendizaje automático hasta el procesamiento del lenguaje natural. Algunas de las bibliotecas más reconocidas son:

NumPy: Es una biblioteca fundamental para la computación científica en Python. Proporciona potentes estructuras de datos, principalmente arreglos multidimensionales (arrays) y una amplia gama de funciones para trabajar con estos arreglos. NumPy es esencial para realizar operaciones numéricas eficientes en Python y es la base de muchas otras bibliotecas en el ecosistema de ciencia de datos y aprendizaje automático.

Pandas: Es una biblioteca muy popular para manipulación y análisis de datos en Python. Ofrece estructuras de datos poderosas y flexibles, principalmente el DataFrame, que permite la manipulación, limpieza y análisis de datos de manera eficiente. Pandas es ampliamente utilizada en tareas como la exploración de datos, la transformación de datos, la agregación y el procesamiento previo antes de aplicar algoritmos de aprendizaje automático. Su facilidad de uso y su integración con otras bibliotecas hacen que sea una herramienta fundamental en el flujo de trabajo de ciencia de datos y análisis de datos en Python.

Scikit-learn (sklearn): Esta biblioteca es una de las principales para el aprendizaje automático en Python. Ofrece una amplia gama de algoritmos de aprendizaje supervisado y no supervisado, así como herramientas para la evaluación y selección de modelos. Sklearn también proporciona utilidades para la selección de características, la validación cruzada y la manipulación de dato

TensorFlow: Desarrollado por Google, TensorFlow es una de las bibliotecas más utilizadas para el aprendizaje profundo y la creación de redes neuronales. Proporciona herramientas para construir y entrenar modelos de IA de manera eficiente.

Keras: Aunque ahora es parte de TensorFlow, Keras se utilizaba como una interfaz de alto nivel para la construcción y entrenamiento de redes neuronales. Es conocido por su facilidad de uso y su enfoque en la rápida experimentación.

[6]

Aprendizaje supervisado y no supervisado

El aprendizaje automático se divide principalmente en dos categorías principales: aprendizaje supervisado y aprendizaje no supervisado.

Aprendizaje Supervisado:

El aprendizaje supervisado implica entrenar un modelo utilizando un conjunto de datos etiquetado. Esto significa que cada ejemplo en el conjunto de datos de entrenamiento está asociado con una etiqueta o resultado deseado. El objetivo es que el modelo aprenda la relación entre las entradas y las salidas esperadas para hacer predicciones precisas en datos nuevos y no vistos.

Hay dos tipos comunes de problemas en el aprendizaje supervisado:

Regresión: Se utiliza cuando la tarea implica predecir un valor numérico o continuo. Por ejemplo, predecir el precio de una casa basado en sus características.

Clasificación: Se utiliza cuando se necesita clasificar los datos en categorías o clases diferentes. Por ejemplo, predecir si un correo electrónico es spam o no spam.

Los algoritmos de aprendizaje supervisado incluyen regresión lineal, regresión logística, máquinas de soporte vectorial (SVM), árboles de decisión, redes neuronales, entre otros.
[7]

Aprendizaje No Supervisado:

En contraste, el aprendizaje no supervisado se utiliza cuando se tienen datos sin etiquetar y el objetivo principal es descubrir patrones, estructuras ocultas o relaciones intrínsecas dentro de los datos.

Los algoritmos de aprendizaje no supervisado se utilizan para:

Agrupamiento (Clustering): Agrupar datos similares en conjuntos o clústeres. Por ejemplo, agrupar clientes según sus hábitos de compra.

Reducción de la dimensionalidad: Reducir la cantidad de características o variables manteniendo la información más relevante. Esto es útil para visualización de datos o para reducir la complejidad en modelos de aprendizaje automático.

Asociación: Encontrar reglas o relaciones de asociación entre diferentes variables. Por ejemplo, en análisis de mercado para identificar patrones de compra asociados.

Los algoritmos comunes de aprendizaje no supervisado incluyen K-Means, Análisis de Componentes Principales (PCA), Algoritmos de Agrupamiento Jerárquico, Reglas de Asociación (como Apriori), entre otros.

[8]

Análisis contrafactual

El contrafactual en el contexto del machine learning se refiere a un enfoque que intenta estimar o predecir el resultado alternativo de una situación o evento particular, considerando lo que habría sucedido si las condiciones o acciones hubieran sido diferentes a las observadas.

En términos más simples, el contrafactual se centra en entender qué habría ocurrido si se hubieran tomado decisiones distintas en el pasado. Este enfoque es crucial en diversas

aplicaciones, como la optimización de políticas, la evaluación de modelos de machine learning en entornos simulados y la comprensión de causalidad en conjuntos de datos.

Por ejemplo, en la evaluación de políticas de intervención en la salud pública, se podría usar el enfoque contrafactual para estimar cómo habría sido la propagación de una enfermedad si se hubieran aplicado diferentes medidas de control.

En el aprendizaje automático, los modelos contrafactuales pueden emplearse para estimar el efecto causal de una intervención o tratamiento, considerando lo que habría sucedido si un individuo o grupo de individuos hubiera recibido un tratamiento diferente al que realmente recibieron.

Estos métodos de contra actualidad suelen ser complejos debido a la necesidad de considerar múltiples escenarios hipotéticos y ajustar los datos de manera apropiada para capturar los efectos causales. Su aplicación correcta puede ayudar a comprender mejor los efectos de las decisiones y acciones pasadas, lo que resulta valioso en áreas donde comprender la causalidad es fundamental.

[9]

Contexto del Trabajo

Además, las principales causas que refleja el estudio sobre la posibilidad de retiro de colaboradores activos se centran en factores como el salario, estado civil y la presencia de hijos. Estos elementos han demostrado ser influyentes en las decisiones de los empleados para permanecer o retirarse de sus roles laborales

La pregunta central que nos planteamos fue la posibilidad de que los colaboradores activos se retiraran y determinar cuáles son las causas más influyentes. En el entorno hospitalario, como ámbito de estudio, se enfrentan desafíos particulares en la retención de su personal. La alta demanda de servicios médicos, el estrés inherente al ambiente de trabajo y la presión por brindar atención de calidad contribuyen a una tasa significativa de

rotación de empleados en este sector. Esta realidad impacta directamente en la continuidad y la calidad de la atención a los pacientes

La problemática de retención de personal en entornos hospitalarios es crucial, ya que la rotación constante afecta no solo la estabilidad del personal, sino también la calidad de la atención médica brindada. La sobrecarga laboral, la presión por mantener altos estándares de atención y el estrés asociado al entorno médico son factores determinantes que influyen en las decisiones de retiro de los colaboradores

La comprensión de estas dinámicas permitió identificar áreas de enfoque específicas para implementar estrategias que mitigaran la rotación del personal. Gracias al modelo desarrollado, se logró crear una estrategia integral que aborda cada una de estas áreas, con el objetivo de generar más estabilidad laboral y mejorar tanto la continuidad de la atención como el bienestar tanto de los empleados como de los pacientes.

Descripción de la Organización

El trabajo realizado como parte de este informe técnico se llevará a cabo en el la Clínica SOMER Este hospital, reconocido en la región por su excelencia en la atención médica, enfrenta desafíos específicos en la retención de su personal sanitario. Con una plantilla diversa de profesionales de la salud, desde médicos hasta personal de enfermería y administrativo, la institución se enfrenta a la constante búsqueda de estrategias efectivas para mantener a su personal altamente capacitado y comprometido

En este contexto, se aplicarán los conocimientos adquiridos en el diplomado sobre machine Learning en entornos de salud. Se emplearán técnicas avanzadas de análisis de datos para identificar las causas subyacentes de la rotación laboral en el hospital. El objetivo principal será desarrollar estrategias específicas de retención adaptadas a las necesidades de cada departamento, utilizando modelos de machine Learning para optimizar la retención del personal.

Se pondrá énfasis en la capacidad de los modelos de machine Learning para identificar patrones y factores predictivos que influyen en la rotación del personal hospitalario. Esto permitirá diseñar estrategias más efectivas y precisas para retener al equipo médico y de apoyo. Además, se buscará crear un entorno laboral que fomente la satisfacción y el compromiso del personal sanitario, integrando de manera inteligente las soluciones provenientes del aprendizaje automático con las necesidades específicas de cada área dentro del hospital.

El análisis detallado de este contexto permitirá la aplicación práctica de los conocimientos adquiridos en el seminario o diplomado, con el objetivo de contribuir a la

mejora continua de la retención del personal en un entorno hospitalario tan exigente como el de Clínica SOMER.

Ética en Machine Learning

En el contexto de abordar la alta tasa de retiros de personal en entornos hospitalarios mediante el uso de técnicas de machine Learning, es crucial considerar la ética en cada paso del proceso. La aplicación ética de estas tecnologías es fundamental para garantizar que los modelos no solo sean precisos y efectivos, sino también justos, equitativos y responsables.

En primer lugar, la recopilación de datos debe realizarse de manera ética y responsable. Es esencial garantizar la privacidad y confidencialidad de la información del personal, así como también asegurar que los datos utilizados no perpetúen sesgos injustos o discriminación.

Durante la construcción y entrenamiento de modelos de machine Learning, se debe tener en cuenta la transparencia y explicabilidad. Es crucial comprender cómo el algoritmo toma decisiones y qué características o variables influyen en esas decisiones. Esto no solo ayuda a identificar posibles sesgos, sino que también permite corregirlos y tomar medidas para mitigar cualquier impacto adverso.

La equidad es otro principio ético esencial. Los modelos deben ser evaluados para asegurar que no favorezcan a un grupo en particular o perpetúen desigualdades existentes. Es importante realizar pruebas rigurosas para identificar y corregir posibles sesgos en el modelo, garantizando así que las decisiones basadas en machine Learning sean justas para todos los empleados, independientemente de su origen, estado civil u otras características personales.

Además, se debe establecer un monitoreo continuo del modelo una vez implementado, para asegurar que siga siendo ético y que no genere consecuencias no deseadas con el tiempo. Esto implica revisar regularmente su desempeño, evaluando su impacto en el personal y en las operaciones hospitalarias.

[10]

Desarrollo e implementación del aprendizaje

A lo largo del seminario de Machine Learning, logramos consolidar ideas claras sobre cómo abordar y obtener resultados efectivos en la construcción de modelos de Machine Learning.

Es crucial destacar que la calidad de los datos es de vital importancia para el éxito de cualquier modelo de Machine Learning. Durante nuestro análisis, pusimos a prueba varios algoritmos, incluyendo CatBoost Classification, Lasso Regression, Elastic Net Regression, Gaussian Naive Bayes y LightGBM Classification. Sin embargo, el que mejor resultado mostró fue XGBoost Classification.

[11]

Tabla 1:

Top	Version	HPO	Algorithm	F1	AUC	Precision	Accuracy	Recall	Hyperparameters	⌵
<input checked="" type="checkbox"/>	2		XGBoost Classification	0.949	0.986	0.961	0.962	0.937		
<input type="checkbox"/>	2		CatBoost Classification	0.923	0.984	0.909	0.941	0.937		
<input type="checkbox"/>	2		LightGBM Classification	0.940	0.983	0.943	0.955	0.937		
<input type="checkbox"/>	2		Gaussian Naive Bayes	0.614	0.765	0.570	0.687	0.667		
<input type="checkbox"/>	2		Elastic Net Regression	0.701	0.830	0.710	0.779	0.692		
<input type="checkbox"/>	2		Lasso Regression	0.709	0.832	0.721	0.786	0.698		

En la imagen se observa la prueba de múltiples algoritmos junto con diversas métricas que han sido fundamentales para evaluar el rendimiento del modelo. Estos parámetros nos han permitido determinar cuál de ellos mostró un comportamiento superior, siendo XGBoost el algoritmo que logró la mejor clasificación.

También nos enfrentamos a retos al determinar qué variables capturar para comprender adecuadamente la dinámica de los retiros. Sin embargo, con las bases establecidas en el seminario y asegurándonos de contar con datos de calidad, pudimos comenzar con solidez y determinar por dónde iniciar para lograr un modelo de Machine Learning exitoso.

[12]

Para analizar y visualizar los datos, nos apoyamos en una herramienta de análisis de datos llamada QlikSense. Esta herramienta fue fundamental para obtener una comprensión

profunda de nuestros conjuntos de datos y facilitó la identificación de las variables más relevantes para el modelo.

Realizamos un proceso similar al ETL (Extract, Transform, Load) para asegurar la calidad y preparación adecuada de los datos antes de su análisis. Esta etapa fue crucial para garantizar que los datos estuvieran limpios, unificados y listos para su procesamiento con el modelo de Machine Learning.

[13]

Una vez aplicado el modelo XGBoost, observamos un alto impacto en la predicción de retiros. Esta potente técnica de clasificación destacó por su capacidad para identificar y ponderar las variables más influyentes en este proceso, logrando una precisión del 96%.

El modelo nos proporcionó una comprensión profunda de las variables relevantes que afectan la retención del personal. Desde el estado civil y la situación familiar hasta otros aspectos laborales y personales, pudimos identificar cómo estas variables influyen significativamente en las tendencias de retiro. Específicamente, notamos que el estado civil y la presencia de hijos son factores cruciales en las decisiones de retiro del personal.

El modelo XGBoost reveló cómo estas variables impactan en las tendencias de retiro, lo que nos permitió desarrollar estrategias más precisas y adaptadas a las necesidades individuales de nuestro equipo.

Además, la capacidad predictiva de XGBoost resultó excepcional al proporcionarnos una visión anticipada y detallada sobre posibles retiros. Esto nos permitió tomar medidas proactivas para retener talento clave en nuestra institución, asegurando una fuerza laboral más estable y comprometida.

[14]

Hablemos de XGBoost

XGBoost, que significa "Xtreme Gradient Boosting", es un algoritmo de aprendizaje automático basado en árboles de decisión. Se destaca por su eficiencia, velocidad y rendimiento excepcional en problemas de clasificación y regresión. Este algoritmo utiliza un enfoque de conjunto (ensemble) que combina múltiples modelos de árboles de decisión más débiles para formar un modelo más sólido y preciso.

La técnica de Gradient Boosting empleada por XGBoost funciona agregando modelos de árboles de decisión de manera secuencial, mejorando iterativamente el rendimiento general del modelo. A diferencia de otros algoritmos, XGBoost utiliza una función de

pérdida personalizada y un enfoque de regularización que controla la complejidad del modelo, minimizando el sobreajuste y mejorando la generalización a datos nuevos.

Mediante la optimización de la función de pérdida y la utilización de árboles poco profundos y robustos, XGBoost puede manejar conjuntos de datos grandes con una alta dimensionalidad y proporcionar predicciones precisas. Su capacidad para identificar patrones complejos en los datos y su excelente capacidad de generalización lo convierten en una elección popular para problemas de clasificación y regresión en diversas aplicaciones de aprendizaje automático.

[15]

¿Por qué QlikSense?

Es principalmente reconocido por su capacidad en análisis visual de datos y Business Intelligence (BI), también ofrece funcionalidades que pueden ser complementarias al ámbito del Machine Learning.

Algunas razones por las cuales QlikSense puede ser útil en el contexto del Machine Learning incluyen:

Preparación de datos avanzada: QlikSense permite a los usuarios realizar preparación avanzada de datos, lo que resulta fundamental en el proceso de Machine Learning. Facilita la integración de datos de múltiples fuentes, la limpieza, transformación y agregación de datos para crear conjuntos de datos preparados para ser utilizados en algoritmos de Machine Learning.

Visualización de resultados del modelo: Una vez que se aplican modelos de Machine Learning, QlikSense puede ayudar a visualizar y presentar los resultados de manera clara y comprensible para los usuarios no técnicos. Esto incluye la creación de gráficos interactivos y tableros (dashboards) que permiten a los usuarios explorar y entender los patrones y predicciones generadas por los modelos.

Análisis exploratorio de datos (EDA): QlikSense ofrece herramientas de análisis exploratorio que son útiles en la fase inicial de comprensión de datos antes de aplicar modelos de Machine Learning. Los usuarios pueden explorar la distribución de datos, identificar correlaciones y tendencias, lo que proporciona una comprensión más profunda de los datos antes de utilizar algoritmos de aprendizaje automático.

Capacidad de integración: QlikSense puede integrarse con otras plataformas y herramientas de Machine Learning, permitiendo una transición más fluida de la preparación de datos a la aplicación de modelos y la presentación de resultado

BI en la inteligencia artificial

Dentro del contexto del Machine Learning (aprendizaje automático), Business Intelligence (BI) se utiliza como un conjunto de herramientas, procesos y tecnologías que facilitan la recolección, procesamiento y análisis de datos con el objetivo de tomar decisiones comerciales informadas y estratégicas.

En el ámbito del Machine Learning, el Business Intelligence a menudo se integra para proporcionar una capa adicional de comprensión y visibilidad de los datos. Esto implica la generación de informes detallados, paneles de control interactivos y visualizaciones avanzadas que permiten a los usuarios, como analistas de datos o líderes empresariales, explorar y comprender los resultados de los modelos de Machine Learning de una manera más accesible y comprensible.

El BI en Machine Learning se utiliza para identificar patrones, tendencias y conocimientos valiosos derivados de los modelos de aprendizaje automático, y luego presentar estos resultados de manera que sean fácilmente interpretables por los stakeholders y tomadores de decisiones. Esto ayuda a las organizaciones a utilizar de manera más efectiva los conocimientos obtenidos de los modelos de Machine Learning en la toma de decisiones estratégicas y operativas.

Métricas

F1-score: El F1-score es una medida que combina la precisión y el recall de un modelo en un solo número. Es útil cuando se tienen clases desequilibradas en los datos. Se calcula como la media armónica de la precisión y el recall y su valor varía entre 0 y 1. Un puntaje alto de F1 indica un buen equilibrio entre precisión y recall.

Accuracy: La precisión es una métrica que indica la proporción de predicciones correctas realizadas por el modelo en relación con el total de predicciones. Se calcula dividiendo el número de predicciones correctas entre el número total de predicciones. Es una medida simple pero útil para modelos con clases balanceadas.

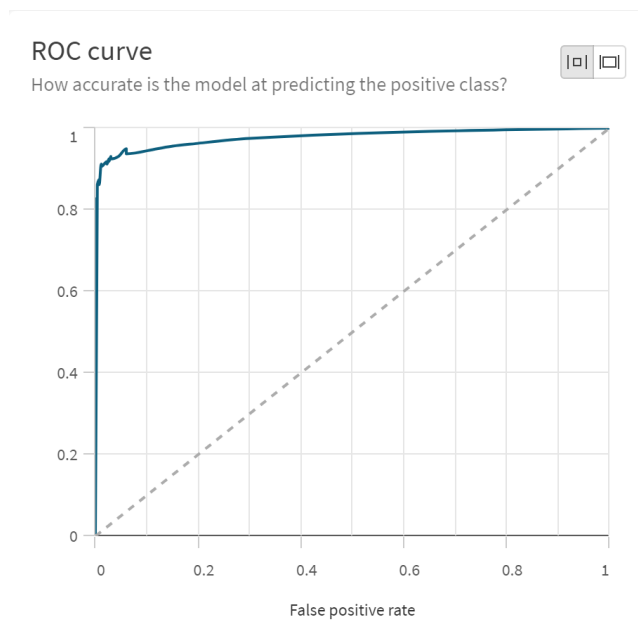
Precision (Precisión): La precisión representa la proporción de instancias positivas identificadas correctamente entre todas las instancias que el modelo predijo como positivas. Se calcula dividiendo el número de verdaderos positivos entre la suma de verdaderos positivos y falsos positivos. Es útil cuando el foco está en minimizar los falsos positivos.

Recall (Sensibilidad o Exhaustividad): El recall mide la proporción de instancias positivas que fueron identificadas correctamente entre todas las instancias positivas reales. Se calcula dividiendo el número de verdaderos positivos entre la suma de

verdaderos positivos y falsos negativos. Es importante cuando se quiere minimizar los falsos negativos.

AUC-ROC (Área bajo la curva ROC): La curva ROC (Receiver Operating Characteristic) muestra la relación entre la tasa de verdaderos positivos (recall) y la tasa de falsos positivos. El AUC-ROC es el área bajo esta curva. Indica la capacidad de un modelo para distinguir entre clases y es útil para evaluar la eficacia del modelo en diferentes umbrales de clasificación. [16]

Figura 1: Curva de ROC con algoritmo XGBoost



Elaborado Por: Elaboración propia a partir de escenario de predicción

La curva ROC representa la relación entre la tasa de verdaderos positivos (Sensibilidad) y la tasa de falsos positivos (1 - Especificidad) a través de distintos umbrales de clasificación. En el eje Y, se muestra la Sensibilidad (TPR - True Positive Rate), que indica la proporción de verdaderos positivos respecto al total de positivos reales. En el eje X, se representa la Especificidad (1 - FPR - False Positive Rate), que refleja la proporción de falsos positivos respecto al total de negativos reales.

La curva ROC visualiza la capacidad de discriminación del modelo a medida que se ajusta el umbral de decisión. Cuanto más se acerca la curva al punto (0,1) del gráfico (esquina superior izquierda), mejor es la capacidad predictiva del modelo. Esto significa que a un umbral más bajo, se logran mayores tasas de verdaderos positivos sin aumentar demasiado los falsos positivos.

[17]

Figura #2: Matrix de confusión para XGBoost

Confusion matrix

How accurate is the model at predicting the target classes?

	Actual true (NO)	Actual false (SI)	
Predicted true (NO)	True positive 144 90.57% <i>Recall</i>	False positive 3 1.13% <i>Fallout</i>	147
Predicted false (SI)	False negative 15 9.43% <i>Miss rate</i>	True negative 263 98.87% <i>Specificity</i>	278
	159	266	Total 425

Elaborado Por: Elaboración propia a partir de escenario de predicción

Esta matriz organiza la información en forma de una tabla, donde las filas representan las clases reales y las columnas las clases predichas por el modelo. En general, la matriz de confusión consta de cuatro secciones:

Verdaderos positivos (VP): Aquí se encuentran los casos en los que el modelo predijo correctamente una muestra como perteneciente a una clase específica, y realmente pertenece a esa clase.

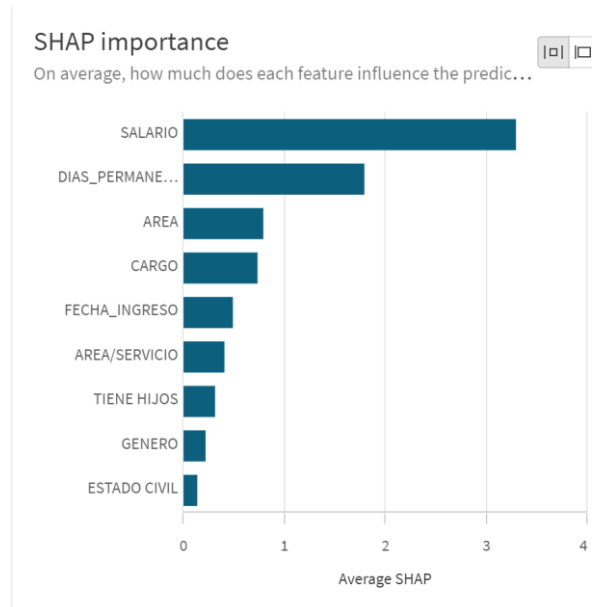
Falsos positivos (FP): Estos son los casos en los que el modelo predijo incorrectamente que una muestra pertenece a una clase específica, cuando en realidad no lo hace.

Falsos negativos (FN): En esta sección se incluyen los casos en los que el modelo predijo incorrectamente que una muestra no pertenece a una clase específica, cuando realmente sí pertenece a esa clase.

Verdaderos negativos (VN): Aquí se encuentran los casos en los que el modelo predijo correctamente que una muestra no pertenece a una clase específica, y realmente no pertenece a esa clase.

[18]

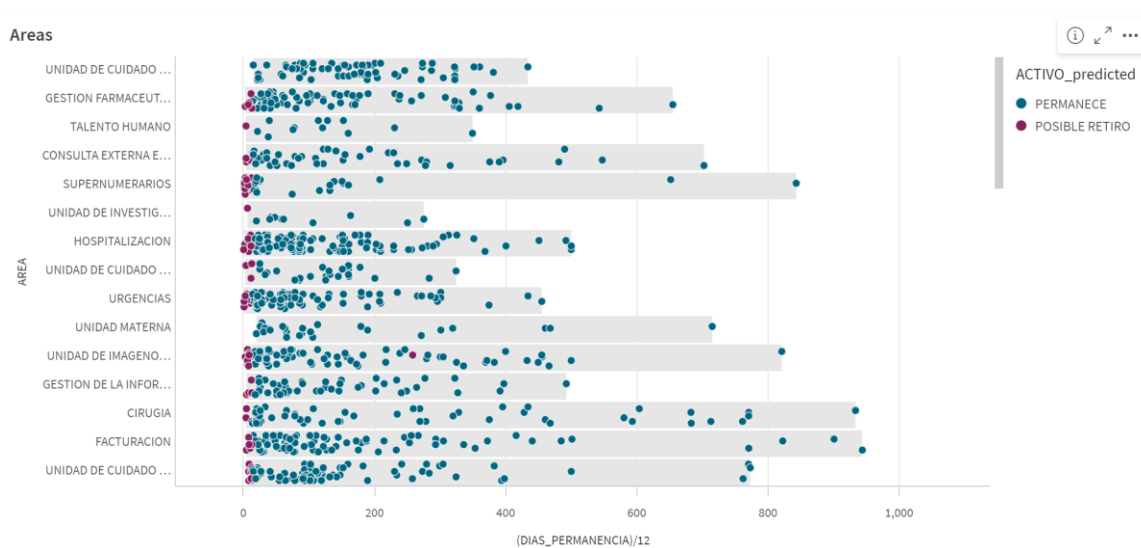
Figura #3: Importancia de factores en algoritmo



Elaborado Por: Elaboración propia a partir de escenario de predicción

SHAP utiliza el concepto de los valores de Shapley de la teoría de juegos para calcular la contribución de cada característica a la predicción del modelo. Estos valores representan cuánto cambia la predicción cuando se agrega una característica específica en comparación con cuando no está presente. Además, SHAP considera todas las combinaciones posibles de características para determinar su impacto en la predicción. [19]

Figura #4: Gráfico de dispersión por áreas con posible retiro

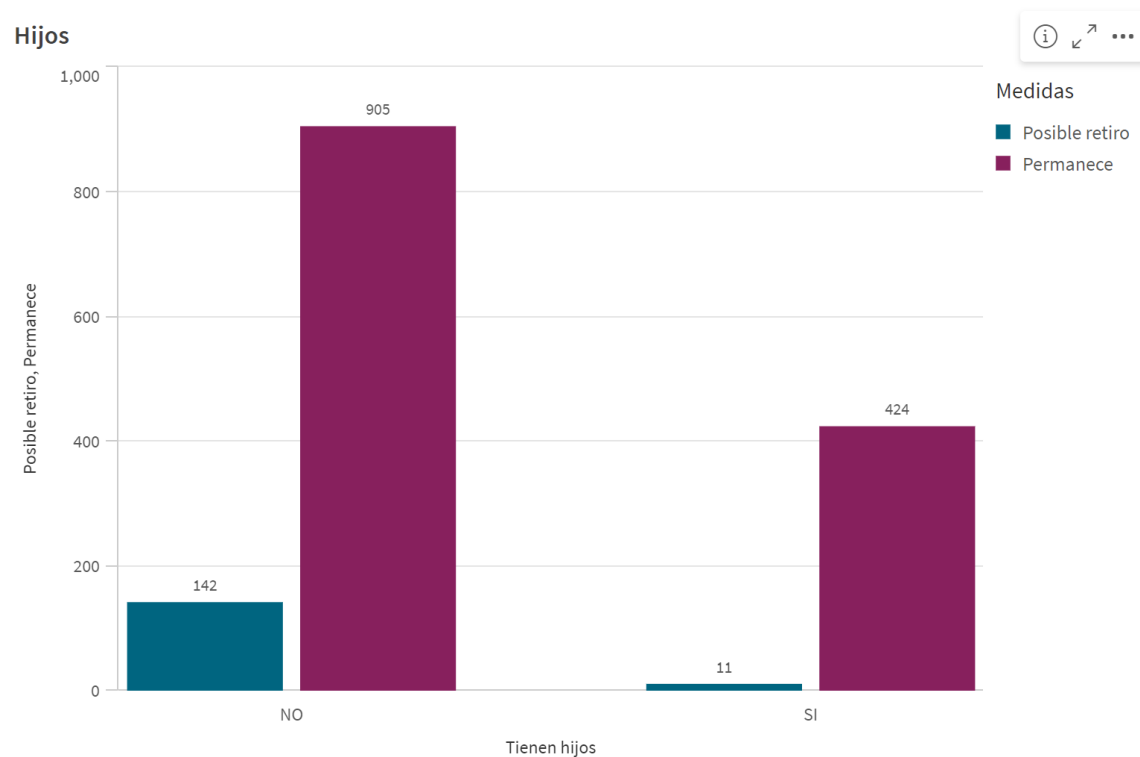


Elaborado Por: Elaboración propia a partir de escenario de predicción

El gráfico de dispersión muestra la relación entre dos variables, representando cada par de datos como un punto en un plano cartesiano.

La disposición de los puntos en el gráfico permite visualizar patrones, tendencias, correlaciones o la ausencia de relaciones entre las dos variables representadas. Algunos patrones comunes que se pueden observar en un diagrama de dispersión [20]

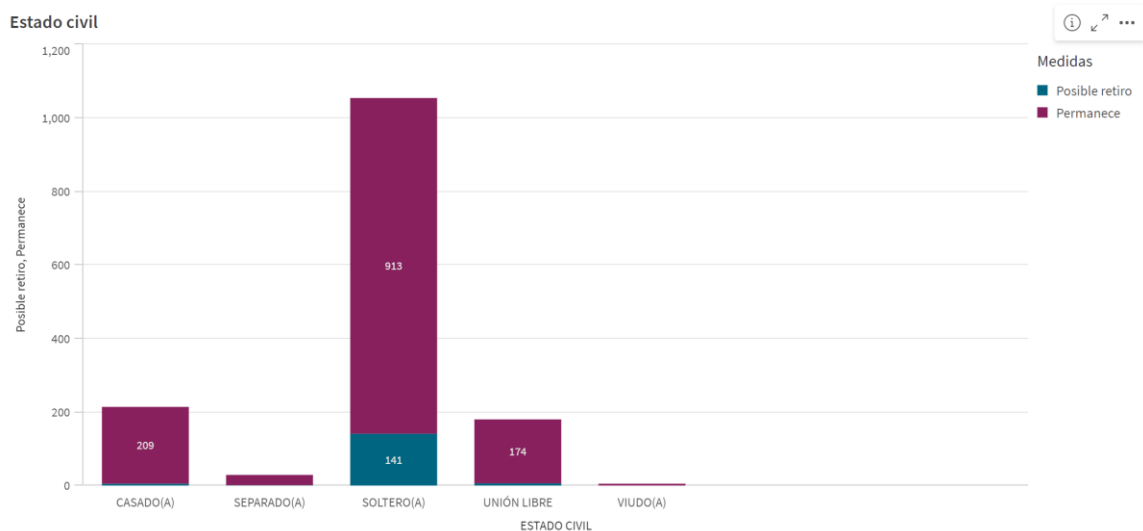
Figura #5: Gráfico de barras por áreas con posible retiro



Elaborado Por: Elaboración propia a partir de escenario de predicción

El gráfico de barras se emplea para mostrar de forma resumida un grupo de datos que puede incluir variables cualitativas y cuantitativas, estos se componen de columnas o barras de diferentes alturas, estas pueden ser horizontales o verticales, Tiene un eje vertical o eje y, donde se ponen los valores que determinan la altura de las barras. A estos números se les conoce como frecuencia. [21]

Figura #6: Gráfico de barras estado civil de colaboradores



Elaborado Por: Elaboración propia a partir de escenario de predicción

El gráfico de barras se emplea para mostrar de forma resumida un grupo de datos que puede incluir variables cualitativas y cuantitativas, estos se componen de columnas o barras de diferentes alturas, estas pueden ser horizontales o verticales, Tiene un eje vertical o eje y, donde se ponen los valores que determinan la altura de las barras. A estos números se les conoce como frecuencia. [22]

Figura #7: Algoritmos y métricas en prueba

	Top	Version	HPO	Algorithm	F1	AUC	Precision	Accuracy	Recall	Hyperparameters
<input checked="" type="checkbox"/>	🔍	2		XGBoost Classification	0.949	0.986	0.961	0.962	0.937	☐
<input type="checkbox"/>		2		CatBoost Classification	0.823	0.984	0.909	0.941	0.937	☐
<input type="checkbox"/>		2		LightGBM Classification	0.940	0.983	0.943	0.955	0.937	☐
<input type="checkbox"/>		2		Gaussian Naive Bayes	0.614	0.765	0.570	0.687	0.667	☐
<input type="checkbox"/>		2		Elastic Net Regression	0.701	0.830	0.710	0.779	0.692	☐
<input type="checkbox"/>		2		Lasso Regression	0.709	0.832	0.721	0.786	0.698	☐

Elaborado Por: Elaboración propia a partir de escenario de predicción

Existen diferentes métricas en machine learning para la evaluación del rendimiento de los modelos, debemos estar en posición de mejorar el poder de predicción general de nuestro modelo antes de que lo ponga en marcha para la producción sobre datos no vistos antes. [23]

Tabla 2.

Base de datos entrevistas de retiro para entrenamiento

A	B	C	D	E	F	G	H	I	J	K	L	M	N
OID	CEDULA	EMPLEADO	GENERO	FECHA_INGRESO	FECHA_RETIRO	DIAS_PERMANENCIA	AREA/SERVICIO	CARGO	SALARIO	AREA	TIENE_HIJOS	ESTADO_CIVIL	ACTIVO
2	39449549	ALEJANDRA I	FEMENINO	7/5/2015			2954 UNIDAD DE GINECOOBSTETRICIA	LIDER UNIDA	5156010	DIRECCION IV SI		CASADO(A)	1
5	15389124	OTONIEL EDL	MASCULINO	9/10/2012			3894 HOSPITALIZACION	AUXILIAR DE	2119344	UNIDAD DE C SI		UNIÓN LIBRE	1
7	43689243	DIANA PATRI	FEMENINO	10/12/2003			7120 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA NO		SOLTERO(A)	1
11	1036927783	LUISA FERNA	FEMENINO	9/12/2010	27/8/2021		3914 ADMINISTRACION	ANALISTA DE	2194328	JURIDICA NO		SOLTERO(A)	0
12	39449580	MARY NELA I	FEMENINO	27/12/2000			8198 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		CASADO(A)	1
13	43714740	LILIANA PATR	FEMENINO	2/6/1998			9137 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		CASADO(A)	1
15	43469536	ANGELA MAF	FEMENINO	5/10/1992			11203 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA NO		SOLTERO(A)	1
16	98709318	JOBAN ANDR	MASCULINO	15/4/2013	21/3/2022		3262 APOYO DIAGNOSTICO	TECNOLOGO	1902827	UNIDAD DE IINO		SOLTERO(A)	0
17	39448175	MARGARITA	FEMENINO	11/2/1998			9248 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		UNIÓN LIBRE	1
18	43714317	GLADIS ESTEL	FEMENINO	16/5/1999			8789 ONCOLOGIA	AUXILIAR DE	2119344	UNIDAD FUNNO		SOLTERO(A)	1
19	21627915	DIANA PAOL	FEMENINO	2/1/2001			8192 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		SEPARADO(A)	1
20	39190202	JASMIN ADRI	FEMENINO	8/2/2011			4503 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA NO		SOLTERO(A)	1
21	43861907	LILIA INES	GUFEMENINO	2/6/2009			5119 ONCOLOGIA	AUXILIAR DE	2119344	UNIDAD FUN SI		CASADO(A)	1
22	1036396907	DANIELA IDA	FEMENINO	9/6/2010			4747 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA NO		SOLTERO(A)	1
23	1036393586	OLGA LUCIA I	FEMENINO	21/1/2008			5617 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		SEPARADO(A)	1
25	56078370	LUZ AYDA MI	FEMENINO	21/4/2008			5526 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		CASADO(A)	1
26	39186047	LUZ ADRIAN	FEMENINO	10/2/1998			9249 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		CASADO(A)	1
27	104129873	DIANA CARO	FEMENINO	14/2/2011	16/12/2022		4323 CIRUGIA - QUIROFANOS	AUXILIAR DE	1718833	CIRUGIA SI		CASADO(A)	0
28	1010174011	DEISY YURAN	FEMENINO	12/5/2009			5140 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		CASADO(A)	1
29	43861569	FLOR ALBA R	FEMENINO	5/8/2003			7247 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		UNIÓN LIBRE	1
31	39192534	SANDRA MILI	FEMENINO	11/5/2004			6967 CIRUGIA - QUIROFANOS	AUXILIAR DE	2119344	CIRUGIA SI		CASADO(A)	1
32	1036940041	LICETH YERAI	FEMENINO	2/2/2021			856 APOYO DIAGNOSTICO	BACTERIOLOGO	4298560	LABORATORI NO		SOLTERO(A)	1
33	43715204	NULL	FEMENINO	13/1/2011			4529 CIRUGIA - QUIROFANOS	AUXILIAR DE	1827104	ESTERILIZACI SI		CASADO(A)	1
34	15437753	NULL	MASCULINO	14/1/2004			7085 CIRUGIA - QUIROFANOS	CAMILLERO	1378933	ENFERMERIA SI		CASADO(A)	1

Tabla 3.

Base de datos entrevistas de retiro para predicción

OID	CEDULA	EMPLEADO	GENERO	FECHA_INGR	FECHA_RETIF	DIAS_PERMA	AREA/SERVIC	CARGO	SALARIO	AREA	TIENE HIJOS	ESTADO CIVI	ACTIVO
4799	1000003235	ANYI KATERI	FEMENINO	4/4/2022			430	HOSPITALIZA AUXILIAR DE	2119344	UNIDAD DE C NO		SOLTERO(A)	
3650	1000100843	JESSICA PAO	FEMENINO	6/12/2021			549	ADMINISTRA AUXILIAR SEI	1583680	GESTION FAF NO		SOLTERO(A)	
4780	1000307419	MABELYN C/	FEMENINO	7/3/2022			458	ADMINISTRA AUXILIAR DE	1591071	TALENTO HU NO		SOLTERO(A)	
3262	1000308423	LAURA CRIST	FEMENINO	5/4/2021			794	CONSULTA E SECRETARIO	1583680	CONSULTA E NO		SOLTERO(A)	
5362	1000308424	LEIDY PAOLA	FEMENINO	7/3/2023			93	HOSPITALIZA AUXILIAR DE	2119344	SUPERNUME NO		SOLTERO(A)	
3611	1000401543	LUISA FERNA	FEMENINO	9/11/2021			576	ADMINISTRA SECRETARIA	1583680	UNIDAD DE I NO		SOLTERO(A)	
2814	1000406843	ANGIE PAOLI	FEMENINO	1/6/2020			1102	HOSPITALIZA AUXILIAR DE	2119344	UNIDAD DE C NO		SOLTERO(A)	
3147	1000414828	SUSANA ALZ	FEMENINO	7/1/2021			882	HOSPITALIZA AUXILIAR DE	2119344	HOSPITALIZA NO		SOLTERO(A)	
5018	1000438837	WILFRANK N	MASCULINO	9/8/2022			303	HOSPITALIZA AUXILIAR DE	2119344	UNIDAD DE C NO		SOLTERO(A)	
5171	1000441210	NATALIA AN	FEMENINO	18/10/2022			233	ONCOLOGIA AUXILIAR DE	2119344	SUPERNUME NO		SOLTERO(A)	
5450	1000539563	MIGUEL DON	MASCULINO	16/5/2023			23	HOSPITALIZA AUXILIAR DE	2119344	SUPERNUME NO		SOLTERO(A)	
5249	1000569311	JUAN CAMIL	MASCULINO	20/12/2022			170	CIRUGIA - QL AUXILIAR DE	2119344	HOSPITALIZA NO		SOLTERO(A)	
5213	1000570478	MARIA CAMI	FEMENINO	22/11/2022			198	HOSPITALIZA AUXILIAR DE	2119344	URGENCIAS NO		SOLTERO(A)	
2672	1000633544	XIMENA GAI	FEMENINO	13/12/2019			1273	UNIDAD DE C AUXILIAR DE	2119344	UNIDAD MA' SI		SOLTERO(A)	
3314	1000640477	JUAN DAVID	MASCULINO	6/5/2021			763	HOSPITALIZA AUXILIAR DE	2119344	URGENCIAS NO		SOLTERO(A)	
1460	1000656618	FRANCI YOH	FEMENINO	18/4/2016			2607	APOYO DIAG SECRETARIA	1470560	UNIDAD DE I NO		SOLTERO(A)	
3382	1000748472	MARIA ALEJ	FEMENINO	1/6/2021			737	HOSPITALIZA AUXILIAR DE	2119344	HOSPITALIZA NO		SOLTERO(A)	
3370	1000751917	ALEJANDRO	MASCULINO	1/6/2021			737	ADMINISTRA AUXILIAR DE	1489032	GESTION DE NO		SOLTERO(A)	
3050	1000756312	MELIZA AGU	FEMENINO	3/11/2020			947	CONSULTA E AUXILIAR DE	2119344	CONSULTA E NO		SOLTERO(A)	
3415	1000760170	VERONICA C	FEMENINO	22/6/2021			716	HOSPITALIZA AUXILIAR DE	2119344	HOSPITALIZA NO		SOLTERO(A)	
5152	1000760933	EMMANUEL	MASCULINO	10/10/2022			241	CIRUGIA - QL AUXILIAR DE	2119344	CIRUGIA NO		SOLTERO(A)	

La columna "activo" se consideró como la variable objetivo o el target en el proceso de modelado predictivo. Esta columna "activo" representa el estado laboral de los empleados, indicando si un empleado posiblemente sigue trabajando en el hospital o si ha dejara su puesto de trabajo

En el contexto del modelado predictivo, se enfocó en predecir la probabilidad o la posibilidad de que un empleado activo se retire. Esta columna "activo" fue el punto focal para el análisis, ya que determinar los factores que influyen en la decisión de retiro del personal activo puede proporcionar información valiosa para implementar estrategias efectivas de retención de empleados.

El objetivo principal era utilizar diferentes variables, como el estado civil, Fechas, días de permanencia, genero, salario, área/servicio, cargo, tiene hijos, para comprender cómo estos factores impactan en la permanencia o la decisión de retirarse del personal activo. El propósito final de este análisis era identificar patrones o tendencias que ayuden a mejorar la retención del personal y, por ende, la calidad y continuidad de la atención médica brindada en el entorno hospitalario.

Conclusiones

Las conclusiones derivadas de este estudio representan un análisis fundamental de las estrategias propuestas para la retención del personal en entornos hospitalarios. La destacada relevancia de estas estrategias reside en su potencial efectividad para reducir de manera significativa las salidas laborales. Estas acciones delineadas no solo podrían contrarrestar la alta rotación de empleados, sino también fortalecer la estabilidad del personal en el ámbito hospitalario.

El estudio demuestra la aplicación práctica de la teoría estudiada durante el seminario, combinada con el ejercicio específico llevado a cabo en el análisis del entorno hospitalario. Un aspecto fundamental fue reconocer la importancia de involucrar personas con conocimientos especializados en el campo de la contratación y la gestión de personal. Esta colaboración permitió comprender mejor el tema y tener en cuenta variables significativas que influyen en la retención del personal hospitalario.

Se ha aprendido valiosamente que no todos los modelos son igualmente efectivos para la predicción, ya que su eficacia depende en gran medida de lo que se desea predecir. Esta revelación resalta la importancia crucial de elegir el enfoque adecuado y los modelos más apropiados para cada escenario específico de predicción en el ámbito de la retención de personal.

Las estrategias identificadas, que incluyen métodos de retención basados en datos históricos, análisis exhaustivos y un enfoque proactivo en factores críticos como el estado civil, la presencia de hijos y otros aspectos sociodemográficos, proporcionan una base sólida para abordar la problemática de la rotación laboral.

La implementación de estas estrategias podría generar un impacto altamente positivo al crear un entorno laboral más satisfactorio y estable, lo que promovería la retención del talento médico esencial para el funcionamiento efectivo de un hospital

Además, para comprender mejor el tema y considerar variables significativas, fue fundamental contar con el apoyo de profesionales con conocimientos en el campo de la

contratación y la gestión de personal. Esta colaboración fue esencial para enriquecer nuestra comprensión del fenómeno de la retención del personal en entornos hospitalarios. El aporte de estos permitió abordar aspectos clave y considerar factores relevantes que influyen en la retención del personal, enriqueciendo así nuestro análisis y las estrategias propuestas para desarrollar el modelo.

Bibliografía

- [1] D. R. Hinestroza, El Machine Learning a través de los tiempos, y los aportes a la humanidad, 2018.
- [2] J. P. Moreno, Una aproximación a Big Data, 2014.
- [3] J. Jiménez, Científico de datos: así es y así se forma uno en esta profesión cada vez más demandada, 2020.
- [4] Coursera, ¿Qué es un ingeniero de datos? Una guía para esta carrera tan demandada, 2023.
- [5] L. A. Lee, M3 - Clase 2: Desmitificando...1, 2021.
- [6] L. A. Lee, M2-Clase 1:Python,Paquetes y librerías, 2021.
- [7] F. J. P.-R. A. M. G.-V. P. G. y. M. J. D. J. C. J. Carmona, Análisis descriptivo mediante aprendizaje supervisado basado en patrones emergentes, 2015.
- [8] L. E. M. Aplicadas, Aprendizaje no supervisado y el algoritmo wake-sleep en redes neuronales, 2012.
- [9] C. tabares, M2 - Clase 1: ¿Qué es un contrafactual? Machine learning, 2021.
- [10] A. C. Orts, Ética de la inteligencia artificial, 2019.
- [11] E. S.-A. M. G.-T. R. M.-C. a. G. I. M. Vázquez-Marrufo, Revisión sistemática de la aplicación de algoritmos de 'machine learning' en la esclerosis múltiple, 2021.
- [12] W. S. Escudero, 2021, Big data y ciencia de datos: conceptos, oportunidades y desafíos.
- [13] S. G. Arboleda, Desarrollo de una herramienta para procesos ETL, 2023.
- [14] V. C. a. F. Fernando, Análisis comparativo de modelos de pronóstico ARIMA y XGBoost aplicados a las series mensuales de ventas en una empresa certificadora, 2021.
- [15] M. S. Sardaña, Modelo predictivo de venta cruzada en productos de vida y salud: Random Forest vs XGBoost, 2022.
- [16] R. Díaz, Métricas de Clasificación.
- [17] L. U. G. y. P. Fernández, Curvas roc, 1998.

- [18] J. M. S. Muñoz, Análisis de Calidad Cartográfica mediante el estudio de la Matriz de Confusión, 2016.
- [19] Qlik, Shap importance.
- [20] J. Castro, Interpretación de los diagramas de dispersión por estudiantes de Bachillerato, 2008.
- [21] G. Global, ¿Qué es un gráfico de barras?.
- [22] Ck-12, Interpretación de gráficos de barras, 2021.
- [23] J. G. G. Ramirez, Métricas De Evaluación De Modelos En El Aprendizaje Automático, 2023.
- [24] D. R. S. y J. S. M. T. Ballestar, ¿Es el big data el siguiente paso en la digitalización de la empresa?, 2018.
- [25] L. A. Lee, M2, 2021.