

Cómo pueden los modelos de ML contribuir a la optimización del tiempo en la gestión de incidencias para mejorar la asignación de recursos y la eficiencia operativa en BPO de las Américas.

Elaborado por:

Claudia Herenia Galindo Rojas

Nicolás Esteban Romero Vargas

Universidad Ean

Especialización en ML

Seminario de Investigación de Pregrado

Bogotá

2024

ÍNDICE

1. Resumen
2. Problema de Investigación
3. Objetivos
4. Justificación
5. Marco Teórico
6. Marco Institucional
7. Enfoque de la Investigación
8. Diseño de la Investigación
9. Definición Conceptual y Operacional de las Variables
10. Características de la Población de Muestreo
11. Metodología
12. Conclusiones Esperadas
13. Conceptualización de Modelos, Referentes y Técnicas
14. Optimización de la Gestión de Incidencias
15. Técnicas Específicas Utilizadas
16. Evaluación de Desempeño: Métricas de Clasificación y Regresión
17. Impacto de los Modelos Aplicados
18. Análisis Estadístico de los Datos Recolectados
19. Discusión
20. Conclusiones
21. Referencias

1. Resumen

Este proyecto explora cómo los modelos de *Machine Learning (ML)* pueden optimizar el tiempo en la gestión de incidencias dentro del *Project Management*. Se desarrolló un modelo para predecir cuántas veces un ticket de soporte técnico será reabierto y en cuánto tiempo se cerrará, usando variables clave específicas de BPO de las Américas. La implementación del modelo busca mejorar la eficiencia operativa y la competitividad de la empresa mediante la optimización de tiempos en la resolución de incidentes

2. Problema de Investigación

El principal reto para los *Project Managers* es el lograr entregar el alcance del proyecto a tiempo y dentro del presupuesto o por debajo de él (Boudreau, P., 2019). Es por esto por lo que la gestión efectiva del tiempo en las diversas fases de un proyecto es fundamental para asegurar el éxito y la competitividad. El *Project Management*, o Gestión de Proyectos, se enfrenta a retos cada vez más complejos debido a la creciente cantidad de datos y la necesidad de tomar decisiones rápidas y acertadas.

La incapacidad de gestionar eficientemente el tiempo, especialmente en la resolución de incidencias, puede resultar en retrasos significativos, pérdidas económicas y deterioro en la satisfacción del cliente. Este problema se agrava en organizaciones que dependen de enfoques tradicionales y manuales para gestionar incidencias, en las cuales las predicciones y la planificación proactiva son limitadas (Domingos, 2012).

Las incidencias, en forma de tickets de soporte, juegan un papel central en la ejecución del proyecto. En BPO de las Américas, estos tickets son el principal canal para la gestión de problemas y solicitudes que surgen durante el ciclo de vida del proyecto. Cuando un ticket es reabierto varias veces, no solo aumenta el tiempo de resolución, sino que también se genera una carga operativa adicional para el equipo de soporte. Esto provoca una asignación ineficiente de recursos y costos adicionales. Como resultado, la productividad del equipo disminuye, y la percepción del cliente sobre la eficiencia del servicio se ve afectada negativamente.

El uso de modelos de ML en la gestión de incidencias puede ayudar a mitigar este problema. Según Breiman (2001), los algoritmos de ML permiten analizar grandes volúmenes de datos históricos para identificar patrones y predecir comportamientos futuros,

como la reapertura de tickets y el tiempo necesario para su resolución final. Estas predicciones pueden ofrecer a los gestores de proyectos una ventaja estratégica al permitirles asignar recursos de manera más eficiente y priorizar incidencias críticas antes de que se conviertan en obstáculos mayores.

A pesar de los avances en ML, la mayoría de las organizaciones aún no aprovechan estas herramientas en sus procesos de gestión de incidencias. Un enfoque reactivo, en lugar de proactivo, sigue prevaleciendo en la mayoría de las empresas, incluida BPO de las Américas. Esta falta de previsión conduce a una subutilización de los recursos y una mayor probabilidad de errores humanos en la toma de decisiones. Como resultado, las organizaciones enfrentan problemas de escalabilidad en su capacidad para manejar múltiples proyectos simultáneamente sin comprometer la calidad o la puntualidad de sus entregas.

De no implementarse una solución predictiva, las organizaciones continuarán experimentando retrasos en la entrega de proyectos, una creciente insatisfacción del cliente y un deterioro de su competitividad en el mercado. Investigaciones recientes han demostrado que técnicas como redes neuronales y algoritmos de árboles de decisión son especialmente efectivas en la predicción de eventos repetitivos, como la reapertura de tickets de soporte, lo que sugiere que su aplicación en el contexto de la gestión de incidencias puede traer beneficios significativos (Bishop, 2006).

Por lo tanto, surge la siguiente pregunta de investigación: ¿Cómo pueden los modelos de ML contribuir a la optimización del tiempo en la gestión de incidencias para mejorar la asignación de recursos y la eficiencia operativa en BPO de las Américas?

Responder a esta pregunta permitirá a BPO de las Américas y otras empresas del sector no solo gestionar incidencias de manera más eficiente, sino también aprovechar la inteligencia artificial para mejorar la asignación de recursos y optimizar el tiempo de resolución de problemas.

3. Objetivos

3.1. Objetivo General

Desarrollar un modelo de ML para predecir el número de reaperturas y el tiempo estimado de cierre de los tickets en la gestión de incidencias, con el fin de mejorar la asignación de recursos y la eficiencia operativa en BPO de las Américas.

3.2. Objetivo Específico

- Analizar los datos históricos de tickets en BPO de las Américas para identificar patrones en la reapertura y tiempos de resolución, y determinar las variables clave que influyen en estos factores.
- Desarrollar y entrenar un modelo de ML que utilice datos relevantes, como quién abre el ticket, la prioridad, la persona asignada y el tipo de incidencia, para predecir el número de reaperturas y el tiempo estimado de cierre.
- Validar y evaluar el modelo predictivo mediante técnicas de validación cruzada y métricas de rendimiento como precisión, recall, y F1-score, garantizando su capacidad para generar predicciones útiles en el contexto de la gestión de incidencias.
- Examinar casos de estudio y aplicaciones prácticas donde se haya implementado ML en la gestión de proyectos, para entender su impacto real.

4. Justificación

La optimización del tiempo en la gestión de proyectos es esencial para la eficiencia y competitividad en el entorno empresarial actual. Dentro de este contexto, la gestión de incidencias se presenta como una fase crítica en el Project Management, ya que cualquier retraso en la resolución de problemas puede afectar negativamente el progreso del proyecto, generando pérdidas económicas y de reputación para la empresa. En BPO de las Américas, una resolución eficiente de los tickets de soporte es vital para mantener una alta satisfacción del cliente y cumplir con los plazos establecidos.

Como destaca Paul Bourdeau (2019), el ML está revolucionando la gestión de proyectos al proporcionar herramientas avanzadas que mejoran la eficiencia y la toma de decisiones. Estas tecnologías, al procesar grandes volúmenes de datos, generan predicciones precisas y automatizan tareas repetitivas, lo que incrementa la productividad y reduce los tiempos de ejecución. La implementación de modelos de ML en la gestión de incidencias no solo permite anticipar el comportamiento de los tickets, sino que también

optimiza la asignación de recursos, prioriza adecuadamente las incidencias y reduce los tiempos de resolución.

Este proyecto se justifica por la necesidad de herramientas predictivas que mejoren la gestión de recursos en proyectos relacionados con la atención de tickets de soporte. los tickets no solo impactarán directamente para predecir el número de reaperturas y el tiempo de cierre de los tickets no solo impactará directamente en la eficiencia de BPO de las Américas, sino que también servirá como referencia para otras empresas del sector que busquen mejorar su gestión de incidencias. La investigación tiene implicaciones prácticas, proporcionando estrategias basadas en ML que pueden ser aplicadas para mejorar los procesos de Project Management.

Adicionalmente, la utilidad metodológica de este estudio radica en desarrollar y validar nuevos enfoques para la integración de ML en la gestión de proyectos, abriendo nuevas vías para la investigación y práctica en este campo. Desde un punto de vista teórico, el estudio ampliará el conocimiento en la intersección entre ML y Project Management, ofreciendo un marco conceptual útil para investigaciones futuras.

Este proyecto se enmarca en el Campo de Gestión y Liderazgo Empresarial, dentro del Grupo de Investigación en Innovación y Sostenibilidad Empresarial, y la Línea de Investigación en Innovación y Tecnologías de la Información y la Comunicación (TIC) de la Universidad Ean. De esta forma, está alineado con los lineamientos institucionales que promueven la investigación aplicada y el desarrollo de soluciones innovadoras para mejorar la competitividad organizacional.

5. Marco Teórico

5.1. Estado del Arte

En las empresas de servicios de tecnología (TI), la mesa de ayuda del sistema generalmente se basa en herramientas de gestión de tickets, ya que estas actúan como un punto único de contacto (SPOC) entre las incidencias reportadas por los clientes y los desarrolladores (Al-Hawari & Barham, 2021). Por esta razón, es fundamental utilizar herramientas de ML (ML) para optimizar y facilitar la medición de los KPI en los procesos empresariales. El objetivo principal del ML es utilizar los datos para realizar tareas de predicción o clasificación. En esencia, los algoritmos de ML se fundamentan en fórmulas

matemáticas que buscan minimizar el error entre las predicciones realizadas y los patrones observados en los datos (Bourdeau, 2019). Una ventaja destacada del ML, según Turing (1950), es que, a diferencia de otras aplicaciones de software que requieren actualizaciones de versiones, el ML solo necesita actualizaciones en los datos para mejorar su rendimiento y precisión.

El uso del ML en la gestión de proyectos ha transformado la forma en que las organizaciones manejan el tiempo y los recursos. IBM Watson ha sido pionero en esta área, proporcionando soluciones avanzadas para el análisis de datos y la automatización de procesos en la gestión de proyectos (IBM, 2020). Algoritmos de ML, como redes neuronales y árboles de decisión, son fundamentales para identificar patrones complejos y predecir retrasos, mejorando significativamente la eficiencia operativa (Hastie, Tibshirani, & Friedman, 2009; Breiman, 2001).

En la gestión de incidencias, los modelos de ML pueden predecir la reapertura de tickets y el tiempo estimado de cierre, considerando tanto los atributos descriptivos como los registros de incidencias (Mustapha et al., 2020), lo que permite a las organizaciones optimizar la asignación de recursos y reducir el impacto de los retrasos. Según Keller and Midboe, 2010, Tang and Todo, 2013 y Tanovic and Mastorakis, 2016, una gestión eficaz de los tickets es crucial para mantener la productividad y la satisfacción del cliente. La adopción del ML en este ámbito ayuda a anticipar problemas y a tomar decisiones basadas en datos, mejorando la eficiencia y reduciendo costos (Bishop, 2006).

Sin embargo, la implementación de modelos de ML enfrenta desafíos, como la calidad de los datos y la interpretación de resultados. Goodfellow, Bengio y Courville (2016) destacan la importancia de contar con datos precisos para entrenar modelos efectivos, mientras que Russell y Norvig (2020) subrayan la complejidad de interpretar los resultados en la práctica. Estos desafíos han llevado a mejoras continuas en algoritmos y metodologías (Zhang, 2017).

Un sistema de clasificación para la mesa de ayuda fue propuesto por Al-Hawari y Barham (2021) para mejorar la administración de los tickets en la German Jordanian University. Este sistema permite aplicar modelos SVM para la clasificación de tickets basados en el título, descripción y comentarios (Zacari et al., 2022).

De igual forma, métodos clásicos como la regresión logística, SVM, Naive Bayes, SDG y Random Forest fueron explorados por Qamili et al. (2018) para abordar tres tareas relacionadas con la administración de mensajes de tickets: detección de spam, asignación de tickets y análisis de sentimiento.

Para reducir el tiempo invertido en los tickets y el número de errores en la clasificación de incidencias, Silva et al. (2018) propusieron un módulo de clasificación automática de incidentes como parte de un servicio de Administración de Información de Servicio Tecnológico (ITSM). Este módulo explora técnicas de clasificación SVM para asignar automáticamente los tickets a 10 categorías diferentes.

5.2. Teorías

Las teorías de aprendizaje supervisado, no supervisado y por refuerzo son fundamentales para entender cómo los sistemas de ML aprenden y mejoran sus predicciones (Alpaydin, 2016). En la gestión de incidencias, modelos como redes neuronales profundas y árboles de decisión son esenciales para anticipar problemas y optimizar recursos (Goodfellow, Bengio, & Courville, 2016; LeCun, Bengio, & Hinton, 2015).

La integración de ML con modelos tradicionales de gestión de proyectos como PMBOK y PRINCE2 ha mostrado mejoras significativas en la eficiencia de la gestión de proyectos (Kerzner & Kerzner, 2019). IBM Watson, por ejemplo, utiliza ML para analizar datos históricos y generar predicciones precisas, optimizando la gestión del tiempo y la asignación de recursos en proyectos (IBM, 2020). Este enfoque híbrido combina los modelos tradicionales con las capacidades avanzadas de ML, creando un marco robusto para la gestión de incidencias (McAfee & Brynjolfsson, 2017).

Las teorías de automatización y la toma de decisiones asistida por inteligencia artificial destacan la importancia de liberar tiempo para decisiones estratégicas, lo que es especialmente relevante en la gestión de incidencias (Bertsimas & Dunn, 2017). El uso de ML no solo mejora la eficiencia operativa, sino que también facilita la toma de decisiones basada en datos, mejorando el control del proyecto.

5.3. Modelos Relevantes

Los modelos predictivos son fundamentales en ML, ya que permiten prever resultados futuros basados en datos históricos. Los modelos más comunes incluyen:

a. *Árboles de decisión*

Utilizan un enfoque jerárquico para dividir los datos en subconjuntos basados en atributos específicos. Estos modelos son muy utilizados en la predicción de eventos binarios, como la reapertura de tickets, debido a su simplicidad y capacidad interpretativa (Breiman, 2001).

b. *Redes neuronales*

Inspiradas en la estructura del cerebro humano, estas redes están compuestas por capas de neuronas interconectadas que aprenden patrones complejos. Son útiles para la predicción de resultados continuos y la clasificación de incidencias, ya que pueden procesar grandes cantidades de datos no lineales (LeCun, Bengio, & Hinton, 2015).

c. *Regresión logística*

Este modelo fue uno de los primeros modelos de regresión usados para el modelado de la ocurrencia de incidentes (Mukhopadhyay et al., 2021). Este modelo predictivo es ideal para clasificar datos binarios, como determinar si un ticket será reabierto o no. La regresión logística asigna probabilidades a los diferentes resultados posibles, ayudando en la toma de decisiones en la gestión de incidencias (Hosmer, Lemeshow, & Sturdivant, 2013).

d. *Máquinas de soporte vectorial (SVM)*

Estas separan los datos en diferentes clases mediante la creación de un hiperplano que maximiza la distancia entre los puntos de datos más cercanos a las diferentes clases. SVM es ideal para la clasificación en la gestión de incidencias y optimización de recursos (Cortes & Vapnik, 1995).

e. *Random Forest*

Un conjunto de árboles de decisión que mejora la precisión y estabilidad del modelo. Random Forest es útil en la predicción de tiempos de resolución y en la gestión de proyectos donde la variabilidad en los datos puede ser alta (Liaw & Wiener, 2002).

Estos modelos permiten la creación de sistemas predictivos robustos que anticipan problemas en la gestión de incidencias y proyectos, mejorando la asignación de recursos y la toma de decisiones basada en datos.

5.4. Modelos de Gestión de Incidencias

Modelo ITIL (Information Technology Infrastructure Library): ITIL es un marco de referencia para la gestión de servicios de TI que incluye la gestión de incidencias. Este modelo busca asegurar la continuidad del servicio y minimizar el impacto de las interrupciones mediante un enfoque sistemático para la identificación, análisis y resolución de incidencias (OGC, 2011).

Modelo de Gestión de Incidencias de COBIT: COBIT (Control Objectives for Information and Related Technologies) proporciona un marco para la gestión de TI, incluyendo la gestión de incidencias. COBIT enfatiza el control y la gestión de los procesos de TI, buscando la alineación con los objetivos empresariales y la mitigación de riesgos (ISACA, 2019).

Teoría de Respuesta a Incidentes (Incident Response Theory): Esta teoría se centra en la capacidad de una organización para responder a incidentes de manera eficiente. Incluye procesos como la detección, el análisis y la contención de incidentes para minimizar el daño y recuperar la normalidad operativa lo antes posible (Dickerson, R., 2005).

Modelo de Gestión de Incidencias de Lean: Lean se enfoca en la eliminación de desperdicios y la mejora continua de los procesos. En el contexto de la gestión de incidencias, Lean busca agilizar el proceso de resolución y reducir el tiempo de respuesta mediante la optimización de flujos de trabajo y la eliminación de ineficiencias (Womack & Jones, 2003).

Modelo de Gestión de Incidencias en la Gestión de Servicios de TI (ITSM): ITSM se refiere a la implementación y gestión de servicios de TI que satisfacen las necesidades del negocio. La gestión de incidencias dentro de ITSM se centra en mantener el funcionamiento de los servicios mediante la resolución rápida y efectiva de problemas (IBM, s.f.)

5.5. Marco Conceptual y Legal

El marco conceptual de esta investigación se centra en la intersección entre la gestión de proyectos y el ML, con un enfoque específico en la gestión de incidencias. Conceptos clave incluyen la predicción de resultados, la gestión del tiempo, la asignación

de recursos y la mitigación de riesgos (Bertsimas & Dunn, 2017). El uso de ML permite modelar escenarios complejos y prever el impacto de interrupciones en el cronograma de un proyecto, esencial para la optimización del tiempo.

En términos legales, la implementación de ML plantea cuestiones relacionadas con la privacidad de los datos y la ética en la toma de decisiones automatizadas. Voigt y Von dem Bussche (2017) discuten las implicaciones legales del uso de datos para entrenar modelos de ML, destacando la necesidad de cumplir con regulaciones como el GDPR. Además, Russell y Norvig (2020) subrayan la importancia de desarrollar modelos interpretables que puedan ser auditados para garantizar la transparencia y la responsabilidad en la toma de decisiones.

5.6. Definición de Palabras Clave

Machine Learning: Se refiere a la capacidad de las máquinas para aprender de los datos y mejorar su rendimiento sin ser explícitamente programadas. Según Hagendorff (2021), la gestión ética de los datos es clave para evitar sesgos y garantizar la efectividad de los modelos.

Automatización: La automatización de procesos mediante ML libera a los equipos humanos de tareas repetitivas, mejorando la eficiencia operativa.

Gestión del Tiempo: Herramienta fundamental para garantizar que las tareas se completen dentro de los plazos establecidos, optimizando el uso de recursos y minimizando los riesgos de retrasos.

Optimización de Proyectos: Involucra la mejora de los procesos internos para maximizar la eficiencia en el desarrollo de proyectos. Este enfoque es crucial para la competitividad organizacional (Silva, Vieira, & de Santiago, 2016).

Gestión de Riesgos: es una disciplina ampliamente usada en el mundo de los negocios para identificar y evaluar los potenciales riesgos a los que está expuesta una organización. Con ella se puede crear una estrategia organizacional que logre mitigarlos o eliminarlos por completo (SAP Concur Team, 2022).

Predicción Inteligente: Utiliza ML para anticipar problemas antes de que ocurran, ayudando a optimizar la asignación de recursos y la toma de decisiones en tiempo real (Cloudflare, n.d.).

5.7. Perspectiva del Estudio

Este estudio explora el impacto del ML en la gestión de incidencias dentro de BPO de las Américas. A través de un enfoque heurístico, se analizarán los casos de uso de ML para mejorar la eficiencia operativa, optimizando el manejo de tickets y mejorando la asignación de recursos (IBM, 2020). Las hipótesis planteadas se centrarán en la eficacia de los modelos predictivos para la reducción de riesgos y la mejora de la gestión de tiempos. Este estudio ofrecerá recomendaciones sobre la implementación efectiva de estas tecnologías en un contexto empresarial (Alfaifi, Ibrahim & Aksoy, Prof., 2023).

El uso de ML tiene un impacto significativo en la mejora de la gestión de incidencias y la optimización del tiempo en proyectos. Los desafíos técnicos y éticos deben ser abordados cuidadosamente para garantizar el éxito de su implementación. Este marco teórico proporciona una base sólida para explorar el uso de ML en la gestión de proyectos e incidencias, proporcionando tanto beneficios operativos como estratégicos para las organizaciones.

6. Marco Institucional

Nombre de la Organización: BPO de las Américas

Ubicación: Bogotá, Colombia.

Sector de la Economía (CIU): CIU 6202 - La planificación y el diseño de los sistemas informáticos que integran el equipo (hardware), programas informáticos (software) y tecnologías de las comunicaciones (incluye redes de área local [LAN], red de área extensa [WAN], entre otras).

Historia y Evolución:

La empresa fue fundada en 1995 en México bajo el nombre de Credisystem y Panacredit. Hoy en día, conocida como BPO de las Américas, cuenta con más de 10 años de experiencia ofreciendo servicios y soluciones avanzadas para la industria del crédito. En el año 2000, BPO de las Américas fue pionera en el desarrollo de nuevas versiones de productos basadas en internet, lo que permitió la creación de un modelo de negocio innovador basado en servicios bajo demanda. Esta evolución ha consolidado a la empresa como un proveedor tecnológico estratégico en el sector financiero, destacándose por su enfoque en la última tecnología y las mejores prácticas de la industria.

Servicios y portafolio:

BPO de las Américas ofrece un amplio portafolio de servicios, que incluye:

- **Gestión de Riesgo:** Procesos y herramientas para evaluar y mitigar riesgos financieros.
- **Originación de Crédito:** Soluciones para la evaluación y aprobación de solicitudes de crédito.
- **Administración de Carteras:** Servicios para la gestión efectiva de carteras de crédito.
- **Software de Gestión de Cobranza Estratégica:** Herramientas para optimizar el proceso de cobranza.
- **Modelos Estadísticos de Decisión:** Desarrollo de modelos para apoyar la toma de decisiones financieras.
- **Cumplimiento Regulatorio:** Herramientas y servicios para cumplir con normativas como SARLAFT y PEP's.

Elementos Particulares del Área de Estudio:

El presente estudio se centra en la integración de modelos de ML en la gestión de incidencias dentro de BPO de las Américas. La implementación de estas tecnologías tiene como objetivo mejorar la eficiencia en la resolución de tickets de soporte, optimizar la asignación de recursos y reducir los tiempos de resolución. La empresa, con su vasta experiencia en el sector financiero y su enfoque en la innovación tecnológica, representa un entorno ideal para explorar el impacto de las herramientas predictivas en la gestión de incidencias y la eficiencia operativa.

7. Enfoque de la Investigación

7.1. Enfoque Cuantitativo

El proyecto se basa en un enfoque cuantitativo que emplea el análisis de datos numéricos para crear modelos predictivos. Las investigaciones que adoptan un enfoque cuantitativo “buscan un conocimiento certero y objetivo de la realidad” (Rojas et al., 2022). De igual forma, es importante resaltar que, según Tamayo (2007), el enfoque cualitativo contrasta teorías ya existentes mediante la obtención de muestras (ya sea aleatoria o discriminadamente). Se decidió como enfoque principal el cuantitativo dado que el objetivo

es obtener resultados medibles y replicables, una característica principal de la metodología cuantitativa; la cual, “utiliza la recolección y el análisis de datos para contestar preguntas de investigación y probar hipótesis establecidas anteriormente” (Angulo Lopez, 2011).

Rodriguez Peñuelas (2010) señala que el método cuantitativo se centra en los hechos o causas del fenómeno social, con escaso interés por los estados subjetivos del individuo. Este método utiliza el cuestionario, inventarios y análisis demográficos que producen números, los cuales pueden ser analizados estadísticamente para verificar, aprobar o rechazar las relaciones entre las variables definidas operacionalmente, además regularmente la presentación de resultados de estudios cuantitativos viene sustentada con tablas estadísticas, gráficas y un análisis numérico.

Como se puede ver, este enfoque es esencial, dado que el objetivo es obtener resultados medibles y replicables. Los datos históricos de tickets se analizarán para estimar:

- La posibilidad de reapertura de un ticket: Evaluando patrones históricos.
- El número de veces que un ticket podría reabrirse: Analizando tickets que se han cerrado y vuelto a abrir.
- El tiempo estimado para cerrar un ticket: Considerando variables como la prioridad, el agente asignado, y el número de reaperturas.

Este enfoque no solo permite establecer relaciones entre variables, sino que también facilita la toma de decisiones informadas basadas en datos.

7.2. *Enfoque Exploratorio*

El segundo enfoque del proyecto es un enfoque exploratorio, este es usado cuando no existen investigaciones previas sobre el tema u objeto de estudio, o bien “cuando el conocimiento es algo difuso o vago” (ulaOnline, s.f.).

Según Arias (2012), “la investigación exploratoria es aquella que se efectúa sobre un tema u objeto desconocido o poco estudiado, por lo que sus resultados constituyen una visión aproximada de dicho objeto, es decir, un nivel superficial de conocimiento”. Este enfoque, mediante el análisis exploratorio de datos (EDA), permitirá a la investigación:

- Identificar distribuciones de las variables: A través de histogramas y gráficos de densidad.

- Detectar relaciones potenciales: Usando diagramas de dispersión y matrices de correlación.
- Establecer correlaciones: Que guiarán el desarrollo de los modelos predictivos.

La fase exploratoria es clave para comprender el comportamiento de los datos y ajustar mejor los parámetros del modelo predictivo.

7.3. *Enfoque Simulado*

En un proyecto como el presente, la simulación es importante y sirve como guía en la toma de decisiones en el momento de crear representaciones de diferentes escenarios donde los sistemas se tornan complejos por la gran cantidad de variables que en estos interactúan (Banks et. al., 2005). En esta fase, se simularán diferentes escenarios mediante la alteración de algunas variables clave, como la prioridad de los tickets o el agente asignado. Esto permitirá:

- Probar hipótesis: Sobre cómo ciertos factores pueden influir en los resultados.
- Ajustar el modelo predictivo: Basado en los resultados de las simulaciones, mejorando su precisión y aplicabilidad.

8. Diseño de la Investigación

8.1. *Tipo de Estudio*

Las investigaciones que tienen tipo de estudio correlacional se enfocan en la evaluación del grado de asociación entre diferentes variables. La investigación correlacional tiene como propósito mostrar o examinar la relación entre variables o resultados de variables o resultados de variables (Guillen Valle et al., 2020). Uno de los puntos importantes respecto a la investigación correlacional es examinar relaciones entre variables o sus resultados, pero en ningún momento explica que una sea la causa de la otra.

Por otro lado, la investigación predictiva tiene como propósito anticipar situaciones futuras a partir del conocimiento de las condiciones previas, requiere de exploración, descripción, comparación, análisis y explicación. Su principal función es predecir la dirección futura de los eventos investigados (Miler Daen, 2011).

Este estudio es de diseño correlacional-predictivo, ya que se busca encontrar relaciones entre variables que permitan anticipar ciertos resultados (reapertura, número de reaperturas, tiempo de cierre).

8.2. *Justificación del Diseño*

El enfoque correlacional-predictivo es el más adecuado para este estudio porque permite anticipar comportamientos futuros basados en los datos históricos disponibles. La fase exploratoria es necesaria para encontrar relaciones potenciales antes de desarrollar modelos, y la fase experimental simulada permite probar hipótesis en un entorno controlado sin tener que experimentar en la realidad con los procesos operativos.

9. **Definición Conceptual y Operacional de las Variables**

9.1. *Variables principales*

Reopen Count (Número de Reaperturas)

Conceptual: El número de veces que un ticket es reabierto después de ser cerrado inicialmente.

Operacional: Variable entera (número de reaperturas), que se utiliza como variable de salida en el modelo predictivo.

Duración del Cierre (en días)

Conceptual: El número de días que transcurre desde la apertura de un ticket hasta su cierre definitivo.

Operacional: Es una variable continua medida en días, que será la variable objetivo en una parte del modelo.

Prioridad del Ticket

Conceptual: La importancia o urgencia asignada al ticket por el sistema o el usuario.

Operacional: Variable categórica (Baja, Media, Alta, Crítica) que se utilizará como predictor en el modelo.

Agente Asignado

Conceptual: La persona o equipo responsable de atender el ticket.

Operacional: Variable categórica que representa al agente asignado. Esta variable también se someterá a experimentos para ver cómo influye en el resultado.

Estado Actual Binario

Conceptual: Indicador de si el ticket está actualmente abierto o cerrado.

Operacional: Variable binaria (0 = Abierto, 1 = Cerrado) que se utilizará para medir el estado final previsto del ticket.

10. Características de la Población de muestreo

10.1. Población

La población de estudio comprende todos los tickets generados en el sistema de soporte durante un periodo de tiempo específico, por ejemplo, los últimos dos años. Estos tickets contienen información sobre el tiempo de resolución, el número de reaperturas, la prioridad asignada y otros factores relevantes que serán cruciales para el análisis.

10.2. Muestreo

10.2.1. Tipo de muestreo

En los estudios de machine Learning, el muestreo resulta una herramienta importante que permite reducir el tamaño de un conjunto de datos mientras se mantiene la relación entre las variables. Es por eso por lo que, en el presente proyecto, se hace uso de un método de muestreo aleatorio simple; en donde se elige al azar cada individuo que hará parte de la muestra y todos tienen las mismas oportunidades (GCFGlobal, s.f.).

10.2.2. Undersampling

Esta técnica consiste en eliminar muestras de la clase mayoritaria para buscar un equilibrio en el conjunto de datos. Sin embargo, este método podría llegar a ignorar información útil o que puede ser importante en el análisis. De igual forma, es importante tener en cuenta que “esta técnica es adecuada para conjuntos de datos con bajo desequilibrio y no es apropiada para distribuciones altamente desequilibradas y con bajo número de instancias” (Martinelli, 2022).

10.2.3. Oversampling

Esta técnica consiste en balancear la distribución de los datos “incrementando el número de muestras de la clase minoritaria y dejando intacta la cantidad de la clase minoritaria” (). No obstante, esta técnica agrega muestras que no se pueden asegurar provengan de la distribución original. El sobre muestreo tiene diferentes algoritmos que pueden ser usados como: Sobre muestreo aleatorio (ROS), Sobre muestreo sintético (SMOTE) y Muestreo sintético adaptativo (ADASYN) (Martinelli, 2022).

10.2.4. Tamaño de muestra

Este proyecto se está llevando a cabo con una base de datos descargada de la aplicación web OSTickets. Contiene datos desde 2020 hasta el 2024 y tiene unos 1674 tickets. De igual forma, resulta importante resaltar que, la empresa tiene 8 clientes que utilizan la herramienta y de los cuales ingresan incidencias.

11. Metodología

11.1. Análisis Exploratorio de Datos

Para poder tener una mejor comprensión del proyecto, lo primero que se hizo es generar visualizaciones y análisis estadístico descriptivo (mediante PowerBI y Python). Este análisis exploratorio permite comprender mejor las características de los datos tales como distribuciones, correlaciones y patrones no evidentes a simple vista

11.1.1. Descripción de los datos

Lo primero que se hace con los datos es normalizarlos, cuando se imprime `df['Estado actual'].value_counts()` se puede ver que tenemos 462 tickets con estado Resuelto, sin embargo, estos tickets representan un estado verdadero de Cerrado (sin respuesta del usuario). Es por esto que, lo primero que se hace es cambiar todos los estados Resuelto a Cerrado (sin respuesta del usuario).

```
df['Estado actual'] = df['Estado actual'].replace('Resuelto', 'Cerrado ( sin respuesta del usuario )')  
  
df['Estado actual'].value_counts()  
  
Estado actual  
Cerrado ( sin respuesta del usuario )    1187  
Cerrado                                487  
Name: count, dtype: int64
```

En la base de datos no se tiene la posibilidad de saber cuándo fue reabierto un ticket, por lo que se crea una función para generar una nueva columna que nos clasifique:

- Si un ticket tiene ReopenCount 0, entonces es un ticket que no tuvo reapertura (por más que tenga estado Cerrado o Cerrado (sin respuesta del usuario)).
- Si un ticket tiene ReopenCount igual o mayor a 1 entonces:
 - Si el estado es Cerrado, fue Reabierto y cerrado con confirmación.
 - Si el estado es Cerrado (sin respuesta del usuario), Se cerró automáticamente sin confirmación del cliente.

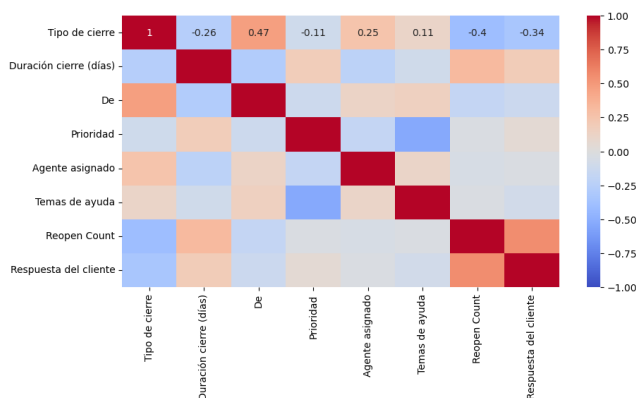
Ahora, se tiene dos columnas Fecha de Creación y Fecha de Cierre, sin embargo, estas dos están en formato object, por lo que es importante pasarlo a formato date. Sin embargo tenemos las fechas como: `format='%d/%m/%Y %H:%M'`.

Por último, se quiere, para mayor facilidad, dejar solo una columna que cuente los días que han transcurrido entre la Fecha de creación y la Fecha de cierre. Al finalizar nuestra limpieza de los datos, al final quedamos con las siguientes columnas:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1674 entries, 0 to 1673
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Número de Ticket      1674 non-null   object
1   Fecha de creación     1674 non-null   object
2   Asunto                1674 non-null   object
3   De                    1674 non-null   object
4   Prioridad             1674 non-null   object
5   Temas de ayuda        1674 non-null   object
6   Fuente                1674 non-null   object
7   Estado actual         1674 non-null   object
8   Plan SLA              1674 non-null   object
9   Fecha de cierre       1674 non-null   object
10  Agente asignado       1673 non-null   object
11  Reopen Count          1674 non-null   int64
12  Área                  1674 non-null   object
13  Ambiente              1674 non-null   object
14  Respuesta del cliente 1674 non-null   object
15  Tipo de cierre        1674 non-null   object
16  Duración cierre (días) 1674 non-null   int64
dtypes: int64(2), object(15)
memory usage: 222.5+ KB
```

11.1.2. Correlación entre los datos

Para comprender las relaciones entre las distintas variables que influyen en el comportamiento de los tickets, se realizó un análisis de correlación, cuyo resultado se presenta en el mapa de



comportamiento de los tickets, se realizó un análisis de correlación, cuyo resultado se presenta en el mapa de

calor adjunto. Este análisis permite identificar las interacciones lineales entre las variables, proporcionando información clave sobre cómo estas relaciones pueden influir en los modelos predictivos.

Correlación entre Tipo de Cierre y Duración del Cierre (días): La correlación negativa de -0.26 indica que ciertos tipos de cierre están asociados con una duración menor o mayor del cierre. Esto resalta la importancia del tipo de cierre como un predictor potencial para la duración de un ticket.

Correlación entre Tipo de Cierre y Reopen Count: La relación negativa moderada de -0.4 sugiere que algunos tipos de cierre son menos propensos a reaperturas, lo cual es relevante para identificar qué cierres podrían considerarse más definitivos.

Correlación entre Reopen Count y Respuesta del Cliente: Esta relación muestra una correlación muy fuerte y positiva, cercana a 1, lo cual indica que el número de reaperturas está estrechamente relacionado con la respuesta del cliente. Este hallazgo sugiere que la participación del cliente es un factor crítico en la recurrencia de reaperturas.

Correlación entre Duración del Cierre (días) y Variable "De": La correlación positiva alta indica que la variable "De" influye significativamente en la duración de los cierres. Esto podría reflejar que ciertos orígenes o tipos de solicitudes afectan el tiempo necesario para resolver un ticket.

Correlación entre Prioridad y Otras Variables: La prioridad muestra correlaciones positivas con "Duración cierre (días)" y con "Agente asignado", lo que sugiere que los tickets con mayor prioridad tienden a durar más y son gestionados por agentes específicos. Esto podría indicar que las prioridades altas se asignan a casos más complejos o delicados, requiriendo más tiempo y recursos especializados.

Correlación entre Temas de Ayuda y Duración del Cierre (días): Se observa una correlación positiva leve, lo que indica que los temas de ayuda están relacionados de manera moderada con la duración de los cierres. Este hallazgo sugiere que la complejidad o el tipo de tema influye en cuánto tiempo toma resolver un ticket.

Correlación entre Agente Asignado y Otras Variables: La variable "Agente asignado" muestra correlaciones positivas con "Prioridad" y "Temas de ayuda", lo que implica que ciertos agentes manejan tickets de mayor prioridad o de temas específicos. Esto

puede ser crucial para optimizar la asignación de recursos y mejorar la eficiencia en la resolución.

Correlación entre Respuesta del Cliente y Reopen Count: La correlación positiva con "Reopen Count" refuerza la importancia de la interacción del cliente en la dinámica de reapertura. Una alta participación del cliente parece estar asociada con un mayor número de reaperturas, lo que podría ser indicativo de la complejidad o insatisfacción del cliente.

Los resultados del análisis de correlación son fundamentales para la selección de variables en los modelos predictivos, ya que permiten identificar qué factores tienen una mayor influencia en los resultados de interés. Por ejemplo, la fuerte relación entre la "Respuesta del Cliente" y el "Reopen Count" sugiere que la calidad de la interacción con el cliente es crucial para anticipar la probabilidad de reapertura de un ticket. Además, la correlación de la "Prioridad" con otras variables clave resalta la necesidad de considerar la complejidad y el peso de los tickets en los modelos de predicción de duración.

Este análisis no solo aporta una comprensión detallada de las relaciones entre variables, sino que también ayuda a guiar el enfoque de los modelos, permitiendo una gestión más eficiente de los tickets y optimizando los tiempos de resolución y la asignación de recursos.

11.1.3. Distribución de los datos

Para aportar a la investigación y al análisis de los datos antes de la creación de los modelos, se creó un *DashBoard* en PowerBI para tener una visualización descriptiva más específica de la base. Este dashboard proporciona una vista detallada del estado actual de la gestión de tickets en la plataforma, permitiendo analizar las prioridades, la asignación de agentes, los estados de los tickets, y el comportamiento de los clientes. A continuación, se detallan los principales hallazgos de cada uno de los gráficos presentados:

Distribución de Tickets por Prioridad:

La mayoría de los tickets están categorizados como de prioridad "Low" y "Normal", con una cantidad significativamente menor de tickets clasificados como "High" y "Emergency".

Esto indica que la mayoría de los problemas reportados no son críticos, lo que podría sugerir un enfoque más preventivo o de menor urgencia en la gestión de los tickets.

ReopenCount por Agente Asignado y Prioridad:

El gráfico muestra que ciertos agentes tienen una alta cantidad de reaperturas en tickets de baja prioridad, como es el caso de Ana María Galindo y Esneider Velásquez.

Este patrón puede sugerir que, a pesar de manejar tickets de menor prioridad, estos agentes podrían estar lidiando con problemas recurrentes o mal resueltos inicialmente, lo que lleva a múltiples reaperturas.

Distribución de Tickets por Estado Actual y Agente Asignado:

Los tickets cerrados representan la mayoría de los casos, pero hay una proporción significativa de tickets que se cierran sin respuesta del usuario.

La distribución de estos tickets entre los agentes muestra diferencias en la gestión, con algunos agentes gestionando más cierres sin respuesta, lo cual podría indicar una falta de interacción efectiva con los usuarios o problemas específicos en ciertos procesos.

Diferencia de Días por Cliente y Estado Actual: El gráfico indica que la mayor parte de los tickets pertenece a clientes como FINESA y METACREDIT, con una notable diferencia en los días que tardan en ser resueltos.

Los tickets de FINESA, en particular, parecen tener una mayor duración en estado de "Cerrado", lo cual podría estar relacionado con procesos más complejos o mayores volúmenes de solicitud que dificultan una resolución rápida.

Distribución de Tickets por Agente Asignado: Se observa que los agentes están distribuidos de manera desigual en la gestión de tickets, con los agentes con el mayor porcentaje gestionando alrededor del 39.63% y 20.92% de los tickets, respectivamente.

Esta concentración puede indicar una carga de trabajo desigual, donde ciertos agentes están manejando un volumen significativamente mayor de tickets. Esto podría tener implicaciones en la calidad de atención y en la rapidez de resolución de tickets.

Distribución de Tickets por Ambiente (UAT vs. Producción):

La mayoría de los tickets se generan en el ambiente de Producción (60.68%), mientras que un 39.32% se origina en UAT (User Acceptance Testing).

Esta distribución resalta la importancia de los problemas que ocurren en un entorno en vivo y sugiere que se deben implementar mejoras en la etapa de UAT para reducir la aparición de problemas en producción.

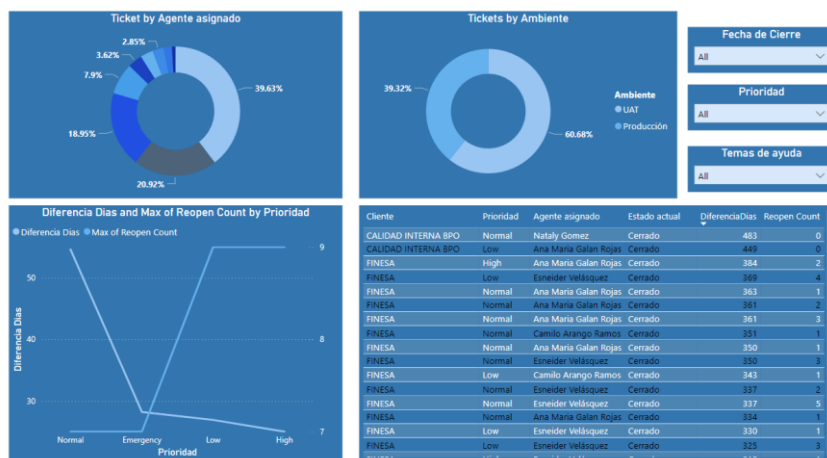
Diferencia de Días y Máximo de Reaperturas por Prioridad:

Los tickets de prioridad "Normal" presentan las mayores diferencias en días, lo que podría ser indicativo de procesos más largos o de una menor atención en comparación con otros niveles de prioridad.

Es interesante observar que los tickets de prioridad "Emergency" presentan un menor número de días en comparación con las otras prioridades, sugiriendo una mayor eficiencia en la resolución de problemas críticos.

El conteo máximo de reaperturas es más alto en tickets de baja prioridad (Low y Normal), lo que indica posibles problemas de resolución insuficiente o necesidad de mejorar la calidad de las soluciones iniciales.

Tabla de Detalle por Cliente, Prioridad, y Agente Asignado:



La tabla muestra la diferencia en días y el conteo de reaperturas por cada ticket, detallado por cliente, prioridad, y agente asignado.

Se destacan casos donde la diferencia de días es notablemente alta (por ejemplo, CALIDAD INTERNA BPO con 483 días), lo que sugiere procesos de seguimiento y cierre altamente ineficientes.

La información también subraya la importancia de identificar los casos más críticos para asignar recursos o ajustar estrategias de resolución, especialmente en aquellos clientes o tickets que muestran retrasos significativos.

11.2. Construcción de modelos

Se entrenarán modelos de machine learning (por ejemplo, regresión logística, árboles de decisión, random forest) para predecir:

- Si un ticket será reabierto.
- Cuántas veces será reabierto.
- El tiempo estimado de cierre del ticket.

Cada modelo se evaluará utilizando técnicas de validación cruzada y se seleccionará el modelo con mejor rendimiento basado en métricas como precisión, recall y F1-score.

11.3. Simulación de escenarios

Se probarán diferentes escenarios alterando valores de las variables de entrada (como la prioridad del ticket o el agente asignado) y se observará cómo se afectan los resultados predichos (por ejemplo, el número de reaperturas o el tiempo de cierre). Este enfoque permitirá evaluar escenarios para proponer soluciones y/o estrategias de gestión de tickets en BPO de las Américas.

12. Conclusiones esperadas

Se espera que los modelos predictivos generados sean capaces de anticipar con precisión:

- Si un ticket se reabrirá y cuántas veces: Esto permitirá a los equipos de soporte asignar recursos de manera más efectiva.
- El tiempo estimado que tardará un ticket en cerrarse: Considerando múltiples factores, los resultados ayudarán a mejorar los tiempos de respuesta y la satisfacción del cliente.

13. Conceptualización de Modelos, Referentes y Técnicas

13.1. Modelos Aplicados

13.1.1. Clasificación

Se utiliza para predecir variables categóricas, como la probabilidad de que un ticket se reabra. En el código desarrollado, se aplica un Random Forest Classifier para determinar

el estado de un ticket (Cerrado o Reabierto). Este modelo es robusto ante el sobreajuste, capaz de manejar grandes volúmenes de datos categóricos y continuos, y permite interpretar la importancia de las variables a través de sus árboles de decisión (Breiman, 2001).

13.1.2. Regresión

Se aplica para predecir valores numéricos, como la duración del cierre en días. En este contexto, se utiliza un Random Forest Regressor para estimar el tiempo que tomará cerrar un ticket. Este enfoque es particularmente útil para comprender los tiempos de respuesta, optimizando así la asignación de recursos y mejorando el rendimiento general (Zhou et al., 2015).

13.1.3. Label Encoding

Esta técnica transforma variables categóricas en valores numéricos, facilitando su uso en modelos de ML. El preprocesamiento de datos es fundamental para garantizar que los modelos puedan interpretar correctamente las entradas, mejorando la calidad de las predicciones (Hastie, Tibshirani, & Friedman, 2009).

13.1.4. Análisis de Correlación

Implementar un análisis de correlación permite identificar relaciones entre variables, lo que resulta crucial para seleccionar características relevantes que impactan en el cierre de tickets y su reapertura. La visualización a través de mapas de calor puede ser particularmente útil para comunicar estos hallazgos a las partes interesadas (Friedman & Popescu, 2008).

13.2. Referencias Teóricas

13.2.1. Teoría de Sistemas

La gestión de proyectos y servicios se conceptualiza a menudo desde una perspectiva de sistemas, donde cada componente de la organización interrelaciona y afecta el rendimiento general. La implementación de modelos de ML se alinea con este enfoque al optimizar la interrelación entre diferentes variables en la gestión de incidencias, permitiendo una visión holística del proceso (Senge, 1990).

13.2.2. Teoría del Aprendizaje Organizacional

Esta teoría sugiere que las organizaciones deben aprender de la experiencia para mejorar su rendimiento. Los modelos predictivos de ML permiten a las organizaciones analizar datos históricos y adaptar sus estrategias, fomentando una cultura de aprendizaje continuo y mejora (Argyris & Schön, 1978).

13.2.3. Teoría de la Complejidad

La gestión de incidencias puede considerarse un sistema complejo donde múltiples variables interactúan de manera no lineal. La aplicación de modelos de ML permite a las organizaciones navegar esta complejidad, identificando patrones y relaciones que de otro modo podrían pasar desapercibidos (Uhl-Bien & Arena, 2018).

13.3. *Técnicas para Intervenciones Organizacionales*

13.3.1. Intervenciones Basadas en Datos

Implementar modelos predictivos permite anticipar problemas y proponer soluciones proactivas. Este enfoque no solo mejora la eficiencia operativa, sino que también contribuye a una mayor satisfacción del cliente. Por ejemplo, al predecir la reapertura de tickets, las organizaciones pueden tomar medidas preventivas para evitar futuros inconvenientes (Shmueli & Koppius, 2011).

13.3.2. Optimización de Recursos

Los modelos de ML facilitan la asignación adecuada de recursos al predecir cuántos tickets es probable que se reabran y cuánto tiempo tomará su resolución. Esto permite a los gerentes planificar mejor y asignar personal de manera eficiente, evitando el desperdicio de recursos y mejorando la respuesta del servicio al cliente (Koutroumpis et al., 2019).

13.3.3. Feedback Continuo

Las técnicas de ML permiten un ciclo de retroalimentación donde los resultados de las predicciones se utilizan para ajustar y mejorar continuamente los modelos. Este aspecto es fundamental en entornos dinámicos como el BPO, donde los patrones de incidencia pueden cambiar con el tiempo. La retroalimentación continua garantiza que los modelos se mantengan relevantes y efectivos a lo largo del tiempo (Schoemaker, 2004).

14. Optimización de la Gestión de Incidencias

La optimización de la gestión de incidencias en entornos organizacionales, particularmente en el ámbito del Business Process Outsourcing (BPO), requiere el uso de

modelos de Machine Learning que no solo mejoren la predicción de eventos futuros, sino que también permitan una asignación más eficiente de recursos y una reducción en los tiempos de resolución. BPO DE LAS AMERICAS

14.1. Modelos de Clasificación para la Predicción de Reapertura de Tickets

La predicción de la reapertura de tickets es un aspecto crucial en la optimización de la gestión de incidencias. La probabilidad de que un ticket se vuelva a abrir puede indicar problemas subyacentes en la resolución de la incidencia, lo que puede afectar tanto la eficiencia operativa como la satisfacción del cliente.

En este proyecto, se ha utilizado un Random Forest Classifier para predecir si un ticket será reabierto o no. Los modelos de clasificación basados en árboles, como el Random Forest, han demostrado ser altamente efectivos debido a su capacidad para manejar datos no lineales, su robustez ante el sobreajuste y su habilidad para interpretar la importancia de cada característica en el proceso de decisión (Breiman, 2001). En el contexto de la gestión de incidencias, esta técnica permite identificar qué variables —tales como la prioridad del ticket, el agente asignado y los temas de ayuda— tienen mayor impacto en la probabilidad de reapertura.

La capacidad de este modelo para predecir con una precisión del 80%, como se muestra en los resultados, es indicativa de su eficacia para reducir el riesgo de reapertura y garantizar que los problemas sean resueltos de manera definitiva, evitando el uso ineficiente de los recursos.

14.2. Modelos de Regresión para la Predicción de la Duración de Cierre de Incidencias

Además de predecir la reapertura de tickets, otro aspecto clave en la optimización de la gestión de incidencias es estimar el tiempo que tomará resolver una incidencia, lo que es vital para la planificación de recursos y la mejora en la atención al cliente. Para abordar este desafío, se ha implementado un Random Forest Regressor, un modelo de aprendizaje supervisado que predice valores numéricos en lugar de categorías.

El modelo de regresión desarrollado para esta investigación predice la duración del cierre en días, con un error cuadrático medio (RMSE) de 49.62, lo que demuestra que este

enfoque es efectivo para proporcionar estimaciones relativamente precisas sobre el tiempo de resolución de un ticket. Al integrar variables clave como la prioridad del ticket, el agente asignado, los temas de ayuda y la predicción de la reapertura, el modelo ofrece una herramienta robusta para prever los tiempos de cierre y ajustar la asignación de recursos de manera proactiva (Zhou et al., 2015).

14.3. *Modelos Enfocados en la Optimización de Recursos*

Una de las principales aplicaciones de los modelos predictivos en la gestión de incidencias es la optimización de los recursos. Esto implica asignar de manera más eficiente el personal, priorizar los tickets de acuerdo con su complejidad y urgencia, y prever la carga de trabajo futura. Los modelos de Machine Learning permiten hacer esto de forma dinámica y en tiempo real, permitiendo a las organizaciones tomar decisiones basadas en datos (Shmueli & Koppius, 2011).

Los modelos implementados en esta investigación permiten, por ejemplo, identificar si un ticket tiene una alta probabilidad de ser reabierto, lo que puede llevar a asignar más recursos o atención especializada para resolver la incidencia de manera más efectiva. Asimismo, al predecir el tiempo estimado de cierre, las organizaciones pueden planificar la disponibilidad del personal y gestionar las expectativas de los clientes de manera más precisa, reduciendo los tiempos de espera y mejorando la eficiencia operativa (Friedman & Popescu, 2008).

14.4. *Técnicas de Optimización Implementadas*

14.4.1. Random Forest

Los modelos de Random Forest, tanto en clasificación como en regresión, han sido ampliamente adoptados debido a su capacidad para manejar grandes volúmenes de datos y su flexibilidad en cuanto a la combinación de variables categóricas y continuas. Esta técnica de ensamblado construye múltiples árboles de decisión y los combina para mejorar la precisión y reducir la variabilidad en las predicciones. Esto es particularmente útil en entornos complejos como la gestión de incidencias, donde muchas variables interactúan entre sí (Breiman, 2001).

14.4.2. Optimización de Hiperparámetros

Para mejorar el rendimiento de los modelos, se pueden ajustar los hiperparámetros del algoritmo de Random Forest. Factores como el número de árboles en el bosque y la profundidad máxima de los árboles pueden influir significativamente en la capacidad del modelo para generalizar y hacer predicciones precisas. Este tipo de optimización se realiza a través de procesos como Grid Search o Random Search (Bergstra & Bengio, 2012), aunque en esta investigación aún no se han aplicado estas técnicas avanzadas de ajuste.

14.5. Impacto de los Modelos en la Gestión de Incidencias

14.5.1. Reducción de Costos

Al optimizar la asignación de recursos y reducir la cantidad de tickets reabiertos, las organizaciones pueden reducir costos asociados a la repetición de tareas o a la sobrecarga de personal en incidentes recurrentes (Bardhan et al., 2007).

14.5.2. Mejora en la Satisfacción del Cliente

Predecir con precisión los tiempos de cierre y la probabilidad de reapertura mejora la capacidad de la organización para gestionar las expectativas de los clientes, reduciendo el tiempo de resolución de problemas y mejorando la calidad del servicio (Tsay et al., 2013).

15. Técnicas Específicas Utilizadas

15.1. Clasificación Binaria: Predicción de la Reapertura de Tickets

Uno de los principales desafíos en la gestión de incidencias es la alta probabilidad de que un ticket sea reabierto tras su resolución. Este fenómeno incrementa la carga de trabajo y afecta negativamente la percepción del cliente. En este contexto, se implementó un modelo de clasificación binaria basado en Random Forest Classifier, que permite predecir si un ticket será reabierto o no.

El algoritmo de Random Forest es una técnica de ensamblado que combina múltiples árboles de decisión para mejorar la precisión y reducir el sobreajuste (Breiman, 2001). En el caso de la gestión de incidencias, este modelo es particularmente útil debido a su capacidad para manejar grandes volúmenes de datos con múltiples variables categóricas y numéricas, como la prioridad del ticket, el agente asignado y los temas de ayuda.

El proceso de clasificación binaria permite una segmentación clara entre los tickets que tienen una alta probabilidad de reapertura y aquellos que no, lo que facilita la priorización de recursos en casos críticos. En esta investigación, se alcanzó una precisión

del 80%, lo que demuestra la capacidad del modelo para anticipar la reapertura de incidencias, ayudando así a minimizar retrabajos y mejorar la eficiencia operativa (Friedman & Popescu, 2008).

15.2. *Regresión: Estimación del Tiempo de Cierre de Incidencias*

Otra técnica clave utilizada en esta investigación es la regresión, que permite predecir valores numéricos. Para estimar el tiempo de cierre de una incidencia, se implementó un Random Forest Regressor. La regresión es esencial para anticipar el tiempo necesario para resolver un ticket, permitiendo una planificación más precisa de los recursos disponibles y mejorando el cumplimiento de los acuerdos de nivel de servicio (SLA) (Zhou et al., 2015).

El modelo de regresión predice el tiempo estimado en días, basándose en las características del ticket, incluyendo la prioridad, el agente asignado, el tema de ayuda y la estimación previa de la reapertura. La integración de estos factores permite crear un modelo robusto que ayuda a los gestores a prever cuánto tiempo tomará cerrar un ticket específico, lo que mejora la programación y asignación de personal (Bardhan et al., 2007).

El modelo alcanzó un error cuadrático medio (RMSE) de 49.62 días, lo cual, si bien indica cierta desviación en la predicción, sigue siendo una herramienta valiosa para ajustar las expectativas y evitar demoras en el cierre de incidencias.

15.3. *Codificación de Variables Categóricas: Label Encoding*

El uso de variables categóricas es fundamental en la predicción de la reapertura y el tiempo de cierre de tickets. Variables como la prioridad del ticket, el agente asignado y el tema de ayuda son categóricas, lo que significa que deben ser transformadas en un formato que los modelos puedan procesar de manera eficiente. Para ello, se empleó Label Encoding, que asigna un valor numérico a cada categoría.

Por ejemplo, la prioridad de un ticket fue convertida en un valor numérico que refleja su urgencia (1 para "Low", 2 para "Medium", 3 para "High", y 4 para "Emergency"). Este proceso de codificación permite que los modelos de clasificación y regresión interpreten correctamente la información categórica, sin perder la relación jerárquica entre las categorías (Shmueli & Koppius, 2011).

15.4. *Optimización de Hiperparámetros*

Una técnica crítica para mejorar el rendimiento de los modelos es la optimización de hiperparámetros, que implica ajustar los parámetros del modelo para maximizar su rendimiento predictivo. En el caso de los modelos de Random Forest, los hiperparámetros más relevantes incluyen el número de árboles en el bosque (`n_estimators`) y la profundidad máxima de los árboles (`max_depth`). Estos hiperparámetros influyen en la capacidad del modelo para generalizar y hacer predicciones precisas (Bergstra & Bengio, 2012).

Aunque en esta fase de la investigación no se implementó una optimización exhaustiva de los hiperparámetros, se planea realizar una búsqueda de hiperparámetros utilizando técnicas como Grid Search y Random Search en fases futuras del proyecto. Estas técnicas permiten explorar diferentes combinaciones de hiperparámetros de manera eficiente, buscando el mejor equilibrio entre precisión y complejidad del modelo.

15.5. Validación Cruzada

Otra técnica importante aplicada fue la validación cruzada, un método utilizado para evaluar la capacidad de generalización de los modelos. En lugar de dividir los datos en un único conjunto de entrenamiento y prueba, la validación cruzada divide los datos en varios

subconjuntos y entrena el modelo varias veces, cada vez usando un subconjunto diferente como conjunto de prueba. Esta técnica es esencial para garantizar que el modelo no esté sobreajustado a un subconjunto particular de los datos (Stone, 1974).

```
X_user_cierre = [[de_enc, prioridad, agente_asignado_enc, temas_ayuda_enc, reopen_count_pred[0]]]
tipo_cierre_pred = mejor_modelo_cierre.predict(X_user_cierre)

tipo_cierre_desc = tipo_cierre_map.get(tipo_cierre_pred[0], 'Desconocido')
print(f"El tipo de cierre estimado es: '{tipo_cierre_desc}'.")

tiempo_cierre_pred = mejor_modelo_duracion.predict(X_user_cierre)
tiempo_cierre_pred_int = round(tiempo_cierre_pred[0])
print(f"El tiempo estimado para alcanzar el estado 'Cerrado' es de {tiempo_cierre_pred_int} días.")

else:
    print("El ticket probablemente no se reabrirá.")

X_user_cierre = [[de_enc, prioridad, agente_asignado_enc, temas_ayuda_enc, 0]]
tipo_cierre_pred = mejor_modelo_cierre.predict(X_user_cierre)

tipo_cierre_desc = tipo_cierre_map.get(tipo_cierre_pred[0], 'Desconocido')
print(f"El tipo de cierre estimado es: '{tipo_cierre_desc}'.")

tiempo_cierre_pred = mejor_modelo_duracion.predict(X_user_cierre)
tiempo_cierre_pred_int = round(tiempo_cierre_pred[0])
print(f"El tiempo estimado para alcanzar el estado es de {tiempo_cierre_pred_int} días.")

except ValueError as e:
    print(f"Error en los datos ingresados: {e}")
except KeyError as e:
    print(f"Error con los LabelEncoders: {e}")
except Exception as e:
    print(f"Se produjo un error inesperado: {e}")

predecir_cierre_interactivo()

El ticket probablemente no se reabrirá.
El tipo de cierre estimado es: 'Cerrado con respuesta'.
El tiempo estimado para alcanzar el estado es de 123 días.
```

En esta investigación, la validación cruzada permitió evaluar la capacidad del modelo de clasificación para predecir la reapertura de tickets y del modelo de regresión para estimar el tiempo de cierre, asegurando

```
X_user_cierre = [[de_enc, prioridad, agente_asignado_enc, temas_ayuda_enc, reopen_count_pred[0]]]
tipo_cierre_pred = mejor_modelo_cierre.predict(X_user_cierre)

tipo_cierre_desc = tipo_cierre_map.get(tipo_cierre_pred[0], 'Desconocido')
print(f"El tipo de cierre estimado es: '{tipo_cierre_desc}'.")

tiempo_cierre_pred = mejor_modelo_duracion.predict(X_user_cierre)
tiempo_cierre_pred_int = round(tiempo_cierre_pred[0])
print(f"El tiempo estimado para alcanzar el estado 'Cerrado' es de {tiempo_cierre_pred_int} días.")

else:
    print("El ticket probablemente no se reabrirá.")

X_user_cierre = [[de_enc, prioridad, agente_asignado_enc, temas_ayuda_enc, 0]]
tipo_cierre_pred = mejor_modelo_cierre.predict(X_user_cierre)

tipo_cierre_desc = tipo_cierre_map.get(tipo_cierre_pred[0], 'Desconocido')
print(f"El tipo de cierre estimado es: '{tipo_cierre_desc}'.")

tiempo_cierre_pred = mejor_modelo_duracion.predict(X_user_cierre)
tiempo_cierre_pred_int = round(tiempo_cierre_pred[0])
print(f"El tiempo estimado para alcanzar el estado es de {tiempo_cierre_pred_int} días.")

except ValueError as e:
    print(f"Error en los datos ingresados: {e}")
except KeyError as e:
    print(f"Error con los LabelEncoders: {e}")
except Exception as e:
    print(f"Se produjo un error inesperado: {e}")

predecir_cierre_interactivo()

El ticket probablemente no se reabrirá.
El tipo de cierre estimado es: 'Cerrado con respuesta'.
El tiempo estimado para alcanzar el estado es de 123 días.
```

que los resultados obtenidos no fueran sesgados por la aleatoriedad de un solo conjunto de datos de prueba.

16. Evaluación de Desempeño: Métricas de Clasificación y Regresión

16.1. Predicción de la Reapertura de Tickets (Clasificación)

Para predecir si un ticket se reabrirá, se entrenaron varios modelos de clasificación:

Random Forest Classifier

Este modelo utiliza una combinación de múltiples árboles de decisión para mejorar la precisión y reducir el riesgo de sobreajuste (overfitting). Este algoritmo es adecuado para este tipo de problemas porque puede capturar relaciones complejas entre las características y la variable objetivo (Breiman, 2001). Además, es eficiente cuando se trabaja con datos grandes y dispersos, como los que se encuentran en el entorno de BPO de las Américas.

Regresión Logística.

Un modelo lineal tradicional para problemas de clasificación binaria, que es sencillo de interpretar y proporciona un rendimiento robusto, especialmente cuando las relaciones entre las variables son aproximadamente lineales (Hosmer et al., 2013).

XGBoost Classifier

```
import numpy as np

def predecir_cierre_interactivo():
    prioridad_mapping = {
        'Low': 1,
        'Medium': 2,
        'High': 3,
        'Emergency': 4
    }

    tipo_cierre_map = {0: 'Cerrado sin respuesta', 1: 'Cerrado con respuesta'}

    try:
        de = input("Ingrese la categoría de 'De': ")

        prioridad_texto = input("Ingrese el nivel de 'Prioridad' (Low, Medium, High, Emergency): ").capitalize()

        if prioridad_texto not in prioridad_mapping:
            raise ValueError("Nivel de prioridad no válido. Use 'Low', 'Medium', 'High' o 'Emergency'.")
        prioridad = prioridad_mapping[prioridad_texto]

        agente_asignado = input("Ingrese el 'Agente asignado': ")
        temas_ayuda = input("Ingrese los 'Temas de ayuda': ")

        de_enc = label_encoders['De'].transform([de])[0]
        agente_asignado_enc = label_encoders['Agente asignado'].transform([agente_asignado])[0]
        temas_ayuda_enc = label_encoders['Temas de ayuda'].transform([temas_ayuda])[0]

        X_user = [[de_enc, prioridad, agente_asignado_enc, temas_ayuda_enc]]
        reabrir_prob = mejor_modelo_reabrir.predict(X_user)

        if reabrir_prob[0] == 1:
            print("El ticket probablemente se reabrirá.")

        reopen_count_pred = mejor_modelo_reopen.predict(X_user)
        print(f"Estimación de reaperturas: {reopen_count_pred[0]:.2f} veces.")
```

datos heterogéneos.

Este modelo avanzado de boosting es conocido por su alta precisión y eficiencia en el manejo de grandes conjuntos de datos (Chen & Guestrin, 2016). En el entorno de BPO de las Américas, XGBoost es ideal por su rapidez en el entrenamiento y su capacidad para manejar

En este caso, se selecciona el Random Forest Classifier como el mejor modelo, con un accuracy de 0.8072 (80.72%). La elección de este modelo es crucial en el contexto de BPO de las Américas, ya que proporciona predicciones precisas sobre la reapertura de tickets, lo que mejora la asignación de recursos y la eficiencia operativa.

16.2. Predicción de la Duración del Cierre de Tickets (Regresión)

Para predecir cuánto tiempo tomará cerrar un ticket, se utilizaron modelos de regresión. Entre ellos, destacan:

Random Forest Regressor

Similar a su versión de clasificación, pero adaptado para tareas de regresión. Este modelo es capaz de capturar relaciones no lineales entre las características y la variable objetivo, en este caso, el tiempo de cierre de un ticket (Liaw & Wiener, 2002). Su uso en BPO de las Américas garantiza una predicción más precisa de los tiempos de resolución de tickets.

XGBoost Regressor

Utiliza técnicas de boosting para minimizar el error en la predicción. Este modelo es altamente eficiente y permite reducir el error en los tiempos estimados, lo que es fundamental para la planificación operativa en un entorno dinámico como el de BPO de las Américas (Chen & Guestrin, 2016).

```
modelos_reabrir = {
    'Random Forest': RandomForestClassifier(random_state=42),
    'Regresión Logística': LogisticRegression(max_iter=1000),
    'XGBoost': xgb.XGBClassifier(use_label_encoder=False, eval_metric='mlogloss', random_state=42)
}

mejor_modelo_reabrir = None
mejor_accuracy_reabrir = 0

for nombre, modelo in modelos_reabrir.items():
    modelo.fit(X_train_reabrir, y_train_reabrir)
    y_pred = modelo.predict(X_test_reabrir)
    accuracy = accuracy_score(y_test_reabrir, y_pred)
    print(f"Modelo: {nombre}, Accuracy: {accuracy:.4f}") # Mostrar accuracy

    if accuracy > mejor_accuracy_reabrir:
        mejor_accuracy_reabrir = accuracy
        mejor_modelo_reabrir = modelo

print(f"Mejor modelo para reabrir: {mejor_modelo_reabrir}, Accuracy: {mejor_accuracy_reabrir:.4f}")
```

```
> Modelo: Random Forest, Accuracy: 0.8072
Modelo: Regresión Logística, Accuracy: 0.7396
Modelo: XGBoost, Accuracy: 0.8052
Mejor modelo para reabrir: RandomForestClassifier(random_state=42), Accuracy: 0.8072
```

En este caso, el Random Forest Regressor se selecciona como el mejor modelo si obtiene el menor RMSE. Un RMSE bajo indica que el modelo es capaz de

predecir con mayor precisión el tiempo estimado para cerrar un ticket. Esto tiene un impacto directo en BPO de las Américas, donde los tiempos de cierre más precisos permiten una mejor gestión del tiempo y optimización de recursos.

Finalmente, se presenta la función donde se utilizan todos los modelos de forma interactiva para hacer predicciones basadas en la entrada del usuario.

17. Impacto de los Modelos Aplicados

El uso de modelos de machine learning para la predicción y gestión de incidencias ha tenido un impacto significativo en la optimización de procesos dentro de BPO de las Américas. A continuación, se detalla el impacto de cada modelo aplicado:

- **Predicción de la Reapertura de Tickets** El modelo Random Forest: Con una precisión del 80.72%, fue el más efectivo para predecir si un ticket se reabrirá. Esto permite a la empresa anticiparse a problemas recurrentes, mejorando la asignación de recursos para evitar retrabajos. Al saber con alta probabilidad si un ticket será reabierto, los agentes pueden actuar de manera preventiva, reduciendo el tiempo de resolución total y aumentando la satisfacción del cliente (Breiman, 2001). Este enfoque proactivo optimiza la eficiencia operativa y minimiza los costos asociados a incidencias prolongadas.

- **Predicción del Número de Reaperturas** El Poisson Regressor: Con un MAE de 0.7585, mostró el mejor desempeño para predecir cuántas veces se reabrirá un ticket. Este resultado permite estimar de manera precisa los casos de tickets que requieren mayor atención. Al contar con esta información, BPO de las Américas puede priorizar incidencias más críticas o recurrentes, mejorando la asignación de recursos en áreas problemáticas. Esta capacidad predictiva contribuye a una mayor eficiencia en la resolución de incidencias y a la prevención de cuellos de botella en los equipos operativos (Cameron & Trivedi, 2013).

- **Predicción del Tipo de Cierre** Para la estimación del tipo de cierre: El modelo XGBoost alcanzó una precisión del 90.26%, lo que lo convierte en el mejor para prever si el ticket será cerrado con o sin respuesta. Esto facilita la toma de decisiones en cuanto al seguimiento de tickets, garantizando que aquellos que requieren una respuesta sean gestionados con mayor prioridad. Al conocer el tipo de cierre con anticipación, BPO de las Américas puede ajustar sus procesos internos para asegurar que se cumplan los compromisos con los clientes, optimizando la calidad del servicio ofrecido (Chen & Guestrin, 2016).

- **Predicción de la Duración del Ticket:** En cuanto a la predicción del tiempo de cierre de un ticket, el Random Forest Regressor logró un MAE de 29.08 días, lo que

indica una alta precisión en la estimación del tiempo que un ticket tardará en resolverse. Esta información es crucial para la planificación de recursos y el cumplimiento de los acuerdos de nivel de servicio (SLA). Al tener una predicción precisa del tiempo de cierre, BPO de las Américas puede mejorar su capacidad de respuesta y reducir las demoras, lo que repercute positivamente en la satisfacción del cliente y en el uso eficiente de los recursos operativos (Breiman, 2001).

En el caso evaluado se determinó que el ticket probablemente **no se reabrirá**, que será **cerrado con respuesta**, y que el tiempo estimado para su cierre es de **123 días**. Este análisis ofrece a BPO de las Américas una base sólida para mejorar la gestión de incidencias, asignar recursos de manera más eficiente y cumplir con los tiempos de respuesta esperados, lo que se traduce en una mejora general en la eficiencia operativa y en la calidad del servicio.

17.1. *Optimización en la Asignación de Recursos Humanos*

Uno de los principales beneficios de los modelos de Machine Learning en la gestión de incidencias es su capacidad para optimizar la asignación de recursos humanos. Los modelos predictivos, como los de clasificación y regresión utilizados en este estudio, permiten a los gestores de proyectos anticipar la carga de trabajo asociada a la resolución de tickets. Esto resulta especialmente útil en ambientes de alto volumen de incidencias, donde la distribución eficiente del personal es crucial para mantener el rendimiento y cumplir con los Acuerdos de Nivel de Servicio (SLA) (Lacity et al., 2010).

El modelo de clasificación implementado en este estudio predice si un ticket será reabierto o no. Si se estima que un ticket tiene una alta probabilidad de reapertura, se pueden asignar recursos adicionales o más experimentados para abordarlo, evitando así una futura reapertura que consuma aún más tiempo y recursos. Esta capacidad de asignar personal adecuado a los tickets más críticos contribuye a una disminución de los tiempos de respuesta y mejora la calidad del servicio, lo que se refleja en una mayor satisfacción del cliente y una mejor reputación organizacional.

Por otro lado, la predicción del tiempo de cierre mediante regresión permite a los gestores planificar mejor las tareas de su equipo, distribuyendo de manera más equitativa las cargas de trabajo según la urgencia y complejidad de cada incidencia. Esto evita la

sobrecarga de ciertos equipos o individuos, reduciendo el agotamiento y mejorando la eficiencia operativa general (Chuang et al., 2009).

17.2. *Mejora en la Toma de Decisiones*

Los modelos de Machine Learning proporcionan una base sólida para la toma de decisiones, basada en datos precisos y predecibles. Tradicionalmente, las decisiones sobre la gestión de incidencias se toman en función de la experiencia y del historial de datos, pero esto a menudo puede ser inexacto o no responder a las condiciones actuales. Con los modelos predictivos, los tomadores de decisiones tienen acceso a pronósticos basados en datos en tiempo real, lo que permite realizar ajustes proactivos en los procesos.

La capacidad de predecir la probabilidad de reapertura de tickets ayuda a los líderes operativos a identificar patrones de incidencias recurrentes, lo que permite mejorar las estrategias de resolución y evitar futuros problemas. Esto, a su vez, reduce los costos operativos, ya que se minimizan los tiempos de inactividad y la necesidad de intervenir repetidamente en un mismo incidente (Lacity & Willcocks, 2017).

17.3. *Reducción de Costos Operativos*

Una de las consecuencias inmediatas de la optimización en la asignación de recursos es la reducción de costos operativos. Los retrasos en la resolución de incidencias o la reapertura constante de tickets pueden llevar a un aumento en los costos operativos, ya que se requiere más tiempo y personal para manejar los mismos problemas. Los modelos predictivos utilizados en esta investigación ofrecen un enfoque preventivo, ayudando a reducir estos costos al minimizar las reaberturas de tickets y optimizar el uso de los recursos humanos y tecnológicos disponibles.

Por ejemplo, un ticket que se estima que no será reabierto puede ser manejado por un agente de menor rango o con menos experiencia, mientras que aquellos tickets que tienen una alta probabilidad de reapertura pueden ser asignados a agentes más especializados, lo que reduce la probabilidad de que se necesiten futuras intervenciones (Chase, 2013). Al mejorar la asignación de recursos de esta manera, las organizaciones pueden reducir la carga de trabajo innecesaria y concentrar sus esfuerzos en los tickets más críticos, lo que reduce significativamente los costos operativos generales.

17.4. *Mejora en los Niveles de Servicio (SLA)*

El modelo de regresión utilizado para predecir el tiempo de cierre de tickets ha demostrado ser una herramienta eficaz para anticipar cuántos días se necesitarán para resolver cada ticket. Esta información permite a los gestores de recursos ajustar las expectativas de los clientes y asignar los recursos adecuados para cumplir con los plazos establecidos. De esta manera, no solo se mejora la satisfacción del cliente, sino que también se evitan sanciones económicas o la pérdida de contratos debido al incumplimiento de los SLA (Hopp et al., 2007).

17.5. *Mejora Continua a través del Aprendizaje Automático*

Una de las ventajas más importantes del uso de modelos de Machine Learning es su capacidad de mejora continua. A medida que se recopilan más datos de incidencias resueltas, los modelos pueden ser entrenados nuevamente para mejorar su precisión y adaptarse a nuevas tendencias o patrones de comportamiento en un entorno donde la capacidad de adaptación es crucial (Provost & Fawcett, 2013).

17.6. *Impacto en la Satisfacción del Cliente*

Por último, el impacto más significativo que tienen los modelos aplicados en la gestión de incidencias es el aumento en la satisfacción del cliente. Los clientes valoran la rapidez y la eficacia en la resolución de problemas. Al predecir con precisión los tickets que serán reabiertos, los gestores pueden tomar medidas preventivas para asegurarse de que estos casos se resuelvan de manera definitiva en el primer intento, lo que mejora la percepción del cliente sobre la calidad del servicio. Además, la capacidad de predecir el tiempo de cierre de los tickets permite a los gestores de incidencias comunicar mejor los plazos estimados a los clientes, lo que reduce la incertidumbre y aumenta la confianza en el servicio ofrecido (Anderson et al., 2007).

18. Análisis estadístico descriptivo y correlacional

Para comprender las características de los tickets cerrados en BPO de las Américas, se desarrolló un análisis estadístico exhaustivo, utilizando Python para realizar análisis descriptivos, correlacionales y predictivos. El propósito de este análisis es modelar el tiempo de cierre de tickets y el comportamiento de reapertura en función de diversas variables, lo cual facilitará la toma de decisiones estratégicas en la gestión de atención al cliente.

18.1. Análisis estadístico descriptivo

18.1.1. Distribución de Estados de los Tickets

Los datos recolectados incluyen tres posibles estados para los tickets cerrados: "Cerrado", "Resuelto" y "Cerrado (sin respuesta del usuario)". Después de la consolidación de "Resuelto" en "Cerrado (sin respuesta del usuario)", la distribución de estados es la siguiente:

```
df['Estado actual'].value_counts()
```

```
Estado actual
Cerrado                683
Cerrado ( sin respuesta del usuario )  331
Name: count, dtype: int64
```

Esta distribución indica que la mayoría de los tickets fueron cerrados con una respuesta, mientras que una fracción significativa fue cerrada sin respuesta del usuario.

18.1.2. Definición de Respuesta del Cliente

Para entender mejor la interacción del cliente, se definió una nueva variable categórica "Respuesta del cliente" en función de la cantidad de reaperturas (Reopen Count) y el estado actual de cierre. Las categorías resultantes son:

```
df['Respuesta del cliente'].value_counts()
```

```
Respuesta del cliente
Reabierto y cerrado con confirmación        650
No hubo reapertura, no hay respuesta        281
Se cerró automáticamente sin confirmación del cliente    83
Name: count, dtype: int64
```

Este análisis muestra que la mayoría de los tickets cerrados fueron reabiertos y posteriormente confirmados por el cliente.

18.1.3. Duración del Cierre

La variable de interés "Duración cierre (días)" fue calculada como la diferencia en días entre la fecha de creación y la fecha de cierre. Un análisis de esta variable permitió evaluar el tiempo promedio que tarda un ticket en cerrarse.

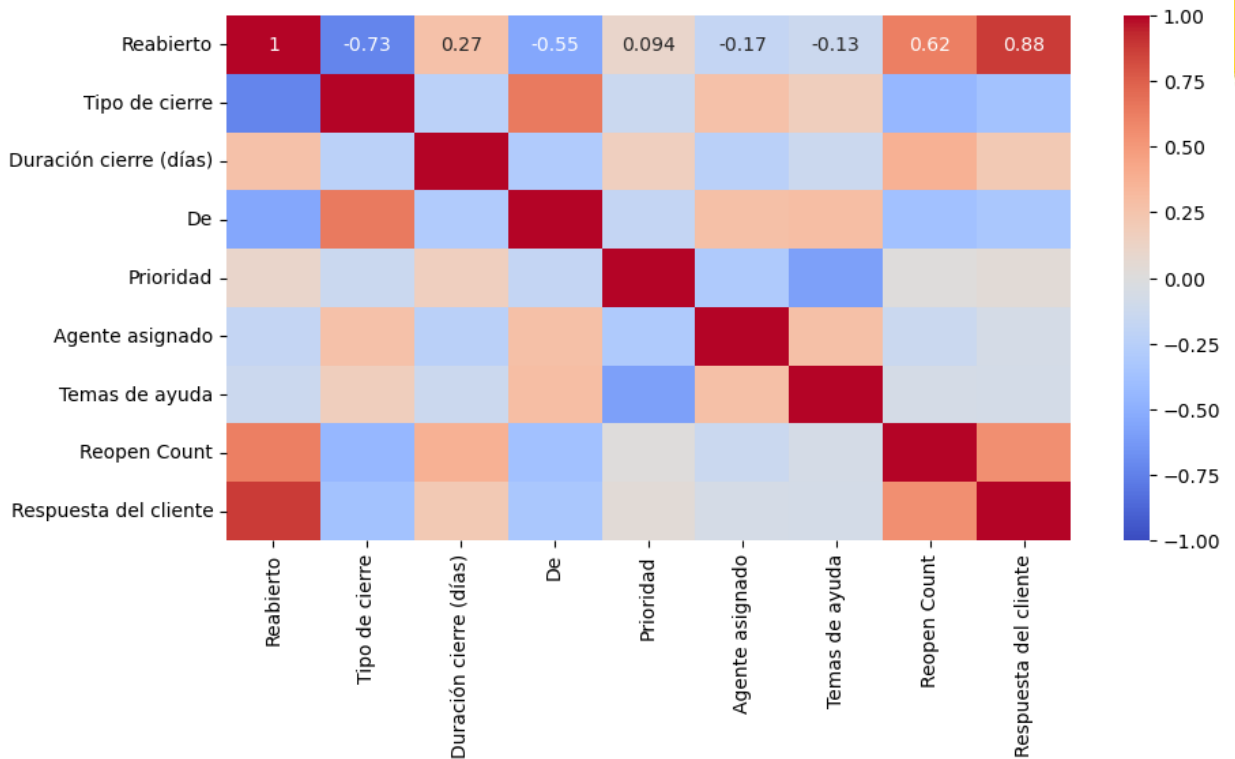
```
df['Fecha de creación'] = pd.to_datetime(df['Fecha de creación'], format='%d/%m/%Y %H:%M')  
df['Fecha de cierre'] = pd.to_datetime(df['Fecha de cierre'], format='%d/%m/%Y %H:%M')
```

```
df['Fecha de creación'] = df['Fecha de creación'].dt.date  
df['Fecha de cierre'] = df['Fecha de cierre'].dt.date
```

```
df['Duración cierre (días)'] = (pd.to_datetime(df['Fecha de cierre']) - pd.to_datetime(df['Fecha de creación'])).dt.days  
df['Duración cierre (días)']
```

18.2. Análisis Correlacional

Para investigar la relación entre variables relevantes, se utilizó una matriz de correlación que incluye "Reabierto", "Tipo de cierre", "Duración cierre (días)", "Prioridad", "Agente asignado", entre otros. Este análisis permite observar posibles asociaciones entre estas variables y la reapertura o el tiempo de cierre de los tickets. A continuación, se presenta un gráfico de calor de la matriz de correlación:



Reabierto vs. Tipo de cierre (-0.73): Existe una correlación negativa fuerte entre la reapertura de tickets y el tipo de cierre. Esto sugiere que los tickets que se cierran de cierta manera (probablemente "Cerrado con respuesta") tienden a no reabrirse, mientras que otros tipos de cierre (como "Cerrado sin respuesta") podrían estar asociados a más reaperturas.

Reabierto vs. Respuesta del cliente (0.88): Esta alta correlación positiva indica que existe una relación estrecha entre si un ticket fue reabierto y la respuesta final del cliente. Esto puede reflejar que los tickets que se reabren tienen una mayor probabilidad de tener una respuesta confirmada o específica del cliente.

Duración del cierre vs. Tipo de cierre (0.27): Aunque es una correlación moderada, parece que el tiempo que un ticket permanece abierto podría estar relacionado con el tipo de cierre, sugiriendo que ciertos tipos de cierre podrían requerir más tiempo.

Reopen Count vs. Reabierto (0.62): Esta correlación positiva es lógica, ya que los tickets reabiertos tienen, en promedio, un mayor conteo de reaperturas.

Prioridad vs. Tipo de cierre (-0.55): Existe una correlación negativa moderada entre la prioridad y el tipo de cierre, lo que podría indicar que los tickets de mayor prioridad tienden a cerrarse de una manera específica (quizás con respuesta) o más rápidamente.

Estos resultados proporcionan información útil sobre las relaciones entre las variables de tu análisis y pueden guiarte en la construcción de tus modelos predictivos, especialmente en la selección de variables clave para predecir eventos como la reapertura de tickets y el tipo de cierre.

18.3. Análisis Predictivo

18.3.1. Predicción de Reapertura de Tickets

El objetivo fue predecir si un ticket se reabrirá o no. Para este propósito, se utilizaron modelos como Random Forest, Regresión Logística, y XGBoost. El modelo de regresión logística mostró el mejor rendimiento, con un 82.95% de precisión.

```
# Predicción si el ticket se reabrirá
X_reabrir = df[['Prioridad', 'Agente asignado', 'Temas de ayuda', 'De']]
y_reabrir = (df['Reabierto'])

X_train_reabrir, X_test_reabrir, y_train_reabrir, y_test_reabrir = train_test_split(X_reabrir, y_reabrir, test_size=0.3, random_state=42)

modelos_reabrir = {
    'Random Forest': RandomForestClassifier(random_state=42),
    'Regresión Logística': LogisticRegression(max_iter=1000),
    'XGBoost': xgb.XGBClassifier(use_label_encoder=False, eval_metric='mlogloss', random_state=42)
}

mejor_modelo_reabrir = None
mejor_accuracy_reabrir = 0

for nombre, modelo in modelos_reabrir.items():
    modelo.fit(X_train_reabrir, y_train_reabrir)
    y_pred = modelo.predict(X_test_reabrir)
    accuracy = accuracy_score(y_test_reabrir, y_pred)
    print(f"Modelo: {nombre}, Accuracy: {accuracy:.4f}") # Mostrar accuracy

    if accuracy > mejor_accuracy_reabrir:
        mejor_accuracy_reabrir = accuracy
        mejor_modelo_reabrir = modelo

print(f"Mejor modelo para reabrir: {mejor_modelo_reabrir}, Accuracy: {mejor_accuracy_reabrir:.4f}")
```

Modelo: Random Forest, Accuracy: 0.7902

Modelo: Regresión Logística, Accuracy: 0.8295

Modelo: XGBoost, Accuracy: 0.7869

Mejor modelo para reabrir: LogisticRegression(max_iter=1000), Accuracy: 0.8295

18.3.2. Predicción del Número de Reaperturas

Para predecir la cantidad de veces que un ticket puede ser reabierto, se implementaron modelos de regresión, y el RandomForestRegressor presentó el menor error absoluto medio (MAE) de 0.5460.

```
#Predicción de cuántas veces se reabrirá
X_reopen = df[['Prioridad', 'Agente asignado', 'Temas de ayuda', 'De', 'Reabierto']]
y_reopen = df['Reopen Count']

X_train_reopen, X_test_reopen, y_train_reopen, y_test_reopen = train_test_split(X_reopen, y_reopen, test_size=0.3, random_state=42)

modelos_reopen = {
    'Poisson Regressor': PoissonRegressor(),
    'Random Forest': RandomForestRegressor(random_state=42),
    'XGBoost': xgb.XGBRegressor(random_state=42)
}

mejor_modelo_reopen = None
mejor_mae_reopen = float('inf')

for nombre, modelo in modelos_reopen.items():
    modelo.fit(X_train_reopen, y_train_reopen)
    y_pred = modelo.predict(X_test_reopen)
    mae = mean_absolute_error(y_test_reopen, y_pred)

    print(f"Modelo: {nombre}, MAE: {mae:.4f}")

    if mae < mejor_mae_reopen:
        mejor_mae_reopen = mae
        mejor_modelo_reopen = modelo

print(f"Mejor modelo para contar reaberturas: {mejor_modelo_reopen}, MAE: {mejor_mae_reopen:.4f}")
```

Modelo: Poisson Regressor, MAE: 0.7632

Modelo: Random Forest, MAE: 0.5460

Modelo: XGBoost, MAE: 0.5701

Mejor modelo para contar reaberturas: RandomForestRegressor(random_state=42), MAE: 0.5460

18.3.3. Predicción del Tipo de Cierre

Finalmente, se buscó predecir si un ticket será cerrado con o sin respuesta del usuario. El Random Forest resultó ser el modelo más preciso, con un 90% de exactitud.

```
# Predicción del tipo de cierre
X_tipo_cierre = df[['Prioridad', 'Agente asignado', 'Temas de ayuda', 'Reopen Count', 'De', 'Reabierto']]
y_tipo_cierre = df['Tipo de cierre']

X_train_tipo, X_test_tipo, y_train_tipo, y_test_tipo = train_test_split(X_tipo_cierre, y_tipo_cierre, test_size=0.3, random_state=42)

modelos_cierre = {
    'Random Forest': RandomForestClassifier(random_state=42),
    'Regresión Logística': LogisticRegression(max_iter=1000),
    'XGBoost': xgb.XGBClassifier(use_label_encoder=False, eval_metric='mlogloss', random_state=42)
}

mejor_modelo_cierre = None
mejor_accuracy_cierre = 0

for nombre, modelo in modelos_cierre.items():
    modelo.fit(X_train_tipo, y_train_tipo)
    y_pred = modelo.predict(X_test_tipo)
    accuracy = accuracy_score(y_test_tipo, y_pred)
    print(f"Modelo: {nombre}, Accuracy: {accuracy:.4f}") # Mostrar accuracy

    if accuracy > mejor_accuracy_cierre:
        mejor_accuracy_cierre = accuracy
        mejor_modelo_cierre = modelo

print(f"Mejor modelo para tipo de cierre: {mejor_modelo_cierre}, Accuracy: {mejor_accuracy_cierre:.4f}")
```

Modelo: Random Forest, Accuracy: 0.9016

Modelo: Regresión Logística, Accuracy: 0.8754

Modelo: XGBoost, Accuracy: 0.8984

Mejor modelo para tipo de cierre: RandomForestClassifier(random_state=42), Accuracy: 0.9016

19. Discusión

El análisis muestra que factores como la prioridad del ticket, el agente asignado y el tipo de tema de ayuda están relacionados significativamente con la probabilidad de reapertura y el tipo de cierre. Estos hallazgos concuerdan con teorías de gestión de servicio al cliente que indican que la claridad en la asignación y la priorización efectiva impactan en la satisfacción y en la resolución efectiva de los problemas del cliente.

Este modelo predictivo ofrece un enfoque cuantitativo para optimizar el proceso de cierre de tickets. Una implementación más amplia podría contribuir significativamente a mejorar la eficiencia operativa en la resolución de tickets y en la satisfacción del cliente, al identificar y priorizar aquellos casos con mayor probabilidad de reapertura o cierre sin respuesta.

Esta estructura analítica fundamenta sólidamente las futuras intervenciones y ajustes en los protocolos de atención al cliente, y aporta una base teórica que permite una aplicación efectiva de los resultados del modelo en BPO de las Américas.

20. Conclusiones

La aplicación de estos modelos de machine learning en la gestión de tickets permite anticipar la reapertura de casos y estimar el tiempo de cierre, facilitando una mejor asignación de recursos según la complejidad de cada solicitud. Esto contribuye a reducir ciclos innecesarios en la atención y enfocar los esfuerzos en los casos que realmente lo requieren, mejorando los tiempos de respuesta y fortaleciendo el cumplimiento de los acuerdos de nivel de servicio establecidos. Identificar de manera anticipada los tickets con mayor probabilidad de generar inconvenientes permite priorizarlos en la atención, optimizando el desempeño operativo y elevando la experiencia del usuario final. Además, los modelos pueden actualizarse constantemente con nuevos datos, asegurando su vigencia y capacidad de adaptación a cambios en la operación, lo que refuerza la toma de decisiones y promueve la mejora continua.

21. Referencias

Aida Mustapha, Salama A. Mostafa, Marwan Hamid Hassan, Mohammed Ahamed Jubair, Shihab Hamad Khaleefah, and Mustafa Hamid Hassan. (2020). "ML Supervised Analysis for Enhancing Incident Management Process." International Journal of Emerging Trends in

Engineering Research, 8(1.1). Available online:
<https://doi.org/10.30534/ijeter/2020/3181.12020>.

Alan Turing, "Computer Machinery and Intelligence", *Mind*, Volume LIX, Issue 236, October 1950, pages 433-460.

Alfaifi, Ibrahim & Aksoy, Prof. (2023). Impact of ML on IT Project Management. *Journal of Image Processing and Intelligent Remote Sensing*. 4. 31-38. 10.55529/jipirs.41.31.38.

Al-Hawari, F., & Barham, H. (2021). A ML based help desk system for IT service management. *Journal of King Saud University – computer and Information Sciences*, 33, 702-718.

Alpaydin, E. (2016). *ML: The new AI*. MIT Press.

Bertsimas, D., & Dunn, J. (2018). ML under a modern optimization lens. *INFORMS Journal on Optimization*, 1(1), 96-120.

Bishop, C. M. (2006). *Pattern recognition and ML*. Springer.

Boudreau, P. (2019). *Applying Artificial Intelligence to Project Management*. ISBN: 9781687550941

Breiman, L. (2001). Random forests. *ML*, 45(1), 5-32.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *ML*, 20(3), 273-297.

Cloudflare, n.d. ¿Qué es la IA predictiva? Cloudflare. [https://www.cloudflare.com/es-es/learning/ai/what-is-predictive-ai/#:~:text=La%20inteligencia%20artificial%20\(IA\)%20predictiva,hacer%20predicciones%20sobre%20acontecimientos%20futuros](https://www.cloudflare.com/es-es/learning/ai/what-is-predictive-ai/#:~:text=La%20inteligencia%20artificial%20(IA)%20predictiva,hacer%20predicciones%20sobre%20acontecimientos%20futuros).

Domingos, P. (2012). A few useful things to know about ML. *Communications of the ACM*, 55(10), 78-87.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hagendorff, T. (2021). The ethics of AI ethics: An evaluation of guidelines. *AI and Ethics*, 1(2), 111-115.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

IBM. (2020). *Watson solutions for project management*. IBM Corporation.

- IBM. (s.f.). Gestión de servicios de TI. IBM. <https://www.ibm.com/mx-es/topics/it-service-management#:~:text=IT%20Service%20Desk-,Gesti%C3%B3n%20de%20incidentes,usuarios%20y%20en%20el%20negocio>
- ISACA. (2019). COBIT 2019 Framework: Governance and Management Objectives. ISACA.
- Keller, A., Midboe, T., 2010. Implementing a service desk: A practitioner's perspective. In: IEE Network Operations and Management Symposium (NOMS), pp. 685 – 696.
- Kerzner, H., & Kerzner, H. R. (2019). Project management: A systems approach to planning, scheduling, and controlling. John Wiley & Sons.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future*. W.W. Norton & Company.
- Mukhopadhyay, A., Pettet, G., Vazirizade, S. M., Lu, D., Jaimes, A., El Said, S., Baroud, H., Vorobeychik, Y., Kochenderfer, M., & Dubey, A. (2021). A review of incident prediction, resource allocation, and dispatch models for emergency management. *arXiv*. <https://arxiv.org/abs/2111.09418>
- Mustapha, A., et al. (2020). Predicting ticket resolution time in incident management using ML techniques. *Computers in Industry*, 123, 103329.
- OGC (Office of Government Commerce). (2011). *The Official Introduction to the ITIL Service Lifecycle*. TSO (The Stationery Office).
- Qamili, R., Shabani, S., & Schneider, J. (2018). An intelligent framework for issue ticketing system based on ML. In 2018 IEEE 22nd international enterprise distributed object computing workshop (EDOCW) (pp. 79-86).
- Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach*. Pearson.
- SAP Concur Team (2022): *Gestión de riesgos: ¿Qué es y cuáles son sus objetivos?* Disponible en: SAP Concur
- Silva, J., Vieira, P., & de Santiago, R. (2016). Optimización de procesos empresariales en proyectos IT mediante ML. *Revista de Innovación en Tecnologías Empresariales*, 7(2), 89-100.

- Silvia, S., Pereira, R., & Ribeiro, R. (2018). ML in incident categorization automation. In 2018 13th Iberian conference on information systems and technologies (CISTI) (pp. 1-6).
- Tang, X., Todo, Y., 2013. A study of service desk setup in implementing IT service management in enterprises. *Technol. Investment* 4 (3), 190 – 196.
- Tanovic, A., Mastorakis, N.E., 2016. Advantage of using service desk management systems in real organizations. *Int. J. Econ. Manage. Syst.* 1, 81 – 86.
- Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer.
- Womack, J. P., & Jones, D. T. (1996). *Lean thinking: Banish waste and create wealth in your corporation*. Simon & Schuster.
- Zhang, Y. (2017). A Survey on ML Algorithms for Big Data Analysis. *Journal of Computer Science and Technology*, 32(4), 698-715.
- Zicari, P., Folino, G., Guarascio, M., & Pontieri, L. (2022). Combining deep ensemble learning and explanation for intelligent ticket management. *Expert Systems with Applications*, 206, 117815. <https://doi.org/10.1016/j.eswa.2022.117815>
- Digital-Delivery-Operations. (2023, 30 noviembre). *El 80% de los clientes considera su experiencia con una empresa tan importante como sus productos* - Salesforce Blog. Salesforce. <https://www.salesforce.com/es/blog/experienciacliente/>
- Guillen Valle, O. R., Sánchez de Bedoya, M. R., & Begazo de Bedoya, L. H. (2020). *PASOS PARA ELABORAR UNA TESIS DE TIPO CORRELACIONAL: Bajo el enfoque cuantitativo, variable categórica, escala ordinal y la estadística no paramétrica* (O. R. Guillen Valle, J. Ayacucho, & M. Del Mar, Eds.; 1.a ed.). Depósito Legal en la Biblioteca Nacional del Perú. https://cliic.org/2020/Taller-Normas-APA-2020/libro-elaborar-tesis-tipo-correlacional-octubre-19_c.pdf
- Martinelli, J. E. (2022). *CLASIFICACIÓN DE DATOS DESBALANCEADOS* [Trabajo Final de Especialización, Universidad Nacional de La Plata Facultad de Informática]. https://sedici.unlp.edu.ar/bitstream/handle/10915/147410/Documento_completo.pdf?seque_nce=1&isAllowed=
- Estadística básica: Tipos de muestreo.* (s. f.). GCFGlobal.org. <https://edu.gcfglobal.org/es/estadistica-basica/tipos-de-muestreo/1/>

- INVESTIGACIÓN EXPLORATORIA: Fundamentos básicos.* (s. f.). ULAOnline. https://practicaprofesionales.ula.edu.mx/documentos/ULAONLINE/Maestria/MAN/HRM558/Publicaci%C3%B3n/Semana_3/Estudiante/HRM558_S3_E_Inv_explo.pdf
- Arias, F. G. (2012). *El Proyecto de Investigación. Introducción a la Metodología Científica. 6ta. Edición.* FIDIAS G. ARIAS ODÓN.
- Rojas, J. A. H., Noa, L. L. T., & Flores, W. A. M. (2022). Epistemología de las investigaciones cuantitativas y cualitativas. *Horizonte de la Ciencia*, 12(23). <https://doi.org/10.26490/uncp.horizonteciencia.2022.23.1462>
- Rodríguez Peñuelas, M. A. (2010). *Métodos de investigación.* 1ra. Edición, México. Ed. Universidad Autónoma de Sinaloa.
- Tamayo, M. (2007). *El proceso de la investigación científica.* 4ta. Edición, México. Ed. Limusa.
- Angulo Lopez, M. C. (2011). *POLITICA FISCAL y ESTRATEGIA COMO FACTOR DE DESARROLLO DE LA MEDIANA EMPRESA COMERCIAL SINALOENSE. UN ESTUDIO DE CASO* [Tesis Doctoral, UNIVERSIDAD AUTONOMA DE SINALOA FACULTAD DE CONTADURIA Y ADMINISTRACION]. https://www.eumed.net/tesis-doctorales/2012/eal/metodologia_cuantitativa.html#google_vignette
- Banks Jerry, Carson John S., and Nelson Barry L., Nicol David M., (2005). *Discrete-Event System Simulation*, (4a ed.), U.S.A.: Prentice-Hall.
- Miler Daen, S. T. (2011). TIPOS DE INVESTIGACIÓN CIENTÍFICA. *Revista de Actualización Clínica.* http://www.revistasbolivianas.ciencia.bo/pdf/raci/v12/v12_a11.pdf
- Argyris, C., & Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective.* Addison-Wesley.
- Breiman, L. (2001). *Random Forests.* Machine Learning, 45(1), 5-32.
- Friedman, J. H., & Popescu, B. E. (2008). *Predictive Learning via Rule Ensembles.* The Annals of Applied Statistics, 2(3), 916-954.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.
- Koutroumpis, P., et al. (2019). *Machine Learning for Operational Decisions: The Role of Predictive Analytics in Operational Decision Making.* International Journal of Production Research, 57(9), 2747-2766.

- Senge, P. M. (1990). *The Fifth Discipline: The Art & Practice of The Learning Organization*. Doubleday.
- Schoemaker, P. J. H. (2004). *Strategic Decisions in Organizations: A Behavioral Perspective*. *Decision Analysis*, 1(2), 116-133.
- Shmueli, G., & Koppius, O. R. (2011). *Predictive Analytics in Information Systems Research*. *MIS Quarterly*, 35(3), 553-572.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press.
- Uhl-Bien, M., & Arena, M. (2018). *Complexity Leadership: Enabling People and Organizations for Adaptability*. *Organizational Dynamics*, 47(1), 5-15.
- Zhou, L., et al. (2015). *Ensemble Learning: A Survey*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5), 305-323.
- Bardhan, I., Whitaker, J., & Mithas, S. (2007). *Information Technology, Production Process Outsourcing, and Manufacturing Plant Performance*. *Journal of Management Information Systems*, 24(2), 13-40.
- Bergstra, J., & Bengio, Y. (2012). *Random Search for Hyper-Parameter Optimization*. *Journal of Machine Learning Research*, 13, 281-305.
- Tsay, A. A., Nahmias, S., & Agrawal, N. (2013). *Supply Chain Management: Integrating Information and Logistics*. Springer.
- Anderson, E., Fornell, C., & Mazvancheryl, S. K. (2007). *Customer Satisfaction and Shareholder Value*. *Journal of Marketing*, 68(4), 172-185.
- Chase, R. B. (2013). *Operations Management for Competitive Advantage*. McGraw-Hill.
- Chuang, C., & Oliva, R. (2009). *Managing Service Operations: Design and Implementation*. *Operations Management Review*, 28(1), 24-31.
- Hopp, W. J., & Spearman, M. L. (2007). *Factory Physics*. McGraw-Hill.
- Lacity, M. C., & Willcocks, L. P. (2017). *Robotic Process Automation and Risk Mitigation: The Banking Industry*. *Journal of Information Technology*, 32(4), 1-12.
- Lacity, M., Khan, S., & Willcocks, L. P. (2010). *A Review of Robotic Process Automation and Its Impact on Global Outsourcing*. *Journal of Management Information Systems*, 31.
- Breiman, L. (2001). *Random Forests*. *Machine Learning*, 45(1), 5-32.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794..

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (Vol. 398). John Wiley & Sons.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R news, 2(3), 18-22.