

Diseño de un modelo de productividad: estrategia para mejorar la eficiencia en ventas del canal presencial de una empresa de telecomunicaciones

Nelson Gómez Muñoz

Asesor

Danitza María Cortes Pérez

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Dedicatoria

A mi esposa, cuyo amor y confianza fueron mi motor en cada paso de este proyecto.

A mi mamá, por enseñarme que con disciplina y humildad se logran los sueños.

A mis amigos más cercanos, por su aliento constante.

Agradecimientos

Quiero expresar agradecimiento a mi asesor de proyecto, Danitza María Cortes Pérez, por su orientación, sus observaciones puntuales y su generosidad al compartir conocimientos.

A los profesores del programa de Especialización en Ciencia de Datos y Analítica de la Universidad Nacional Abierta y a Distancia (UNAD), por brindar sólidas bases académicas y sesiones de formación enriquecedoras que aportaron significativamente al desarrollo de este trabajo.

Agradezco a mi jefe y al gerente de la empresa por permitirme desarrollar este proyecto aplicado, así como por la confianza depositada, su colaboración constante y el valioso intercambio de ideas que enriquecieron significativamente esta experiencia.

Resumen

Este proyecto desarrolló un modelo de aprendizaje automático basado en árboles de decisión para predecir la productividad de los asesores de canal presencial en una empresa de telecomunicaciones. La metodología comprendió cuatro fases: (1) definición de variables clave; (2) recopilación y preprocesamiento de datos; (3) entrenamiento y validación del modelo; y (4) diseño de estrategias de optimización. El modelo final alcanzó una puntuación F1 macro del 74 % y una precisión general del 82 %. Con base en estos resultados, se implementó un plan de acción que incluye un dashboard para seguimiento y programas de capacitación e incentivos específicos para impulsar la productividad. El modelo garantiza una mejora continua, lo que facilita la toma de decisiones estratégicas basadas en datos.

***Palabras clave:** Productividad, Machine Learning, Árbol de decisión, Telecomunicaciones, predicciones.*

Abstract

This project developed a decision-tree-based machine learning model to predict the productivity of in-person channel advisors at a telecommunications company. The methodology comprised four phases: (1) defining key variables; (2) collecting and preprocessing data; (3) training and validating the model; and (4) designing optimization strategies. The final model achieved a macro F1-score of 74% and an overall accuracy of 82%. Based on these results, an action plan was implemented, including a real-time monitoring dashboard and targeted training and incentive programs to boost productivity. The model also establishes a continuous improvement cycle, enabling data-driven strategic decision-making.

Keywords: Productivity, Machine Learning, Decision Tree, Telecommunications, Predictions.

Tabla de Contenido

Introducción	9
Descripción del Problema	10
Planteamiento del Problema.....	10
Pregunta Problema	10
Justificación	11
Objetivos	12
Objetivo General	12
Objetivos Específicos	12
Marco Teórico.....	13
Antecedentes	13
Marco Conceptual.....	16
Marco Legal	26
Desarrollo.....	28
Fase 1 Definición de las Variables Clave	28
Fase 2 Recolección y Preparación de Datos	30
Fase 3 Desarrollo y Evaluación del Modelo	34
Fase 4 Estrategias de Optimización	37
Conclusiones	42
Recomendaciones	43
Bibliografía	44

Lista de Tablas

Tabla 1 <i>Descripción de las Variables</i>	24
Tabla 2 <i>Justificación de las Variables</i>	29
Tabla 3 <i>Balanceo de Datos</i>	34
Tabla 4 <i>Explicación de los Hiperparámetros</i>	35
Tabla 5 <i>Métricas de Evaluación del Modelo</i>	36
Tabla 6 <i>Importancia de las Variables en el Modelo</i>	39
Tabla 7 <i>Estructura de los Incentivos</i>	40

Lista de Figuras

Figura 1 <i>Tipos de Modelos</i>	17
Figura 2 <i>Regresión Lineal Simple</i>	18
Figura 3 <i>Clúster K-Means</i>	19
Figura 4 <i>Estructura de un Árbol de Decisión</i>	21
Figura 5 <i>Árbol de Decisión Artistas al 1 de Billboard</i>	22
Figura 6 <i>Ejemplo de Flujo de Trabajo con Nodos de Bases de Datos</i>	31

Introducción

El canal de ventas presencial de una empresa de telecomunicaciones es una de las áreas más importantes de la empresa, ya que es el pilar fundamental para el cumplimiento de los objetivos comerciales por la venta de servicios, estos canales cuentan con puntos físicos de atención y ventas a nivel nacional por regulación de la superintendencia de industria y comercio, el canal está compuesto por asesores de servicio y venta, supervisores y coordinadores.

La medición de la productividad de los equipos de ventas en la actualidad es un reto debido a la fuerte competencia y las nuevas tecnologías, también al enfoque que muchas empresas tienen y la miden basándose en el desarrollo del asesor y no realizan la medición sobre la gestión de cada asesor, la dificultad para hacer seguimiento en tiempo real e implementar planes de acción para cada uno de los asesores que lo necesiten.

Este proyecto tiene como objetivo desarrollar un modelo de machine learning basado en árboles de decisión que permita evaluar y predecir la productividad en el canal presencial de ventas, analizando variables operativas y demográficas clave. Para ello, se siguió una metodología en cuatro fases: (1) definición de la productividad y selección de variables; (2) recolección y preprocesamiento de datos con técnicas de limpieza y balanceo; (3) entrenamiento, ajuste y validación del modelo; y (4) propuesta de estrategias de intervención guiadas por los resultados.

Con un enfoque basado en datos históricos y métricas de desempeño, este proyecto aporta una herramienta predictiva capaz de predecir y segmentar a los asesores en categorías de productividad.

Descripción del Problema

Planteamiento del Problema

Los canales de ventas presenciales de las empresas de telecomunicaciones son un pilar fundamental para el cumplimiento de sus objetivos comerciales. Sin embargo, enfrentan desafíos constantes para medir la productividad de sus equipos de ventas. La dificultad para realizar un seguimiento en tiempo real y oportuno se traduce en casos de baja productividad y en una reacción tardía para ejecutar planes de acción durante el mes. Además, la demora en la obtención de resultados impide proporcionar feedback a tiempo, lo que dificulta la toma de decisiones estratégicas en la compañía.

Pregunta Problema

¿Cómo puede un modelo de machine learning medir de manera precisa la productividad en el canal presencial de ventas de una empresa de telecomunicaciones, permitiendo un seguimiento en tiempo real y facilitando la toma de decisiones estratégicas?

Justificación

El proyecto aplicado tiene un desafío para el sector de las telecomunicaciones, a falta de herramientas precisas para medir la productividad en el canal de ventas presencial las mediciones actuales no son las más precisas y adecuadas, este problema no permite identificar factores que mejoran o limitan el rendimiento de la fuerza comercial, esto a su vez impacta la toma de decisiones estratégicas.

El principal beneficio con el desarrollo del proyecto aplicado es la creación de un modelo de machine learning que va a permitir medir de manera objetiva y precisa la productividad de la fuerza de ventas del canal presencial, el modelo va a proporcionar información importante para identificar variables clave que impulsan el desempeño, permitirá una optimización en la asignación de los recursos de la empresa, se podrá diseñar estrategias personalizadas que permitirán mejorar la eficiencia operativa identificando patrones y tendencias permitiendo ajustar estrategias de manera dinámica y en tiempo real.

El proyecto aplicado contribuye a ofrecer una solución tecnológica basada en machine learning realizando análisis a datos históricos y actuales identificando variables específicas como características generacionales, métricas de desempeño e incentivos, el modelo no solo cuantifica la productividad, sino que también proporciona información para tomar decisiones. Es importante desarrollar un modelo de productividad basado en machine learning, aunque actualmente se usen modelos tradicionales que suelen ser basados en indicadores comerciales y en algunos casos financieros, incorporar estas tecnologías en la medición de la productividad permitirá a la empresa mantenerse a la vanguardia del mercado optimizando sus recursos, siendo rentables y alcanzando sus objetivos comerciales de manera más efectiva.

Objetivos

Objetivo General

Desarrollar un modelo de machine learning para evaluar y predecir la productividad en el canal de ventas presencial en una empresa de telecomunicaciones, mediante el análisis de variables clave y técnicas avanzadas de aprendizaje automático que optimice la toma de decisiones estratégicas.

Objetivos Específicos

Definir la productividad y las variables clave que impactan en el canal de ventas presencial, desde una perspectiva integral de la operación para establecer una base sólida en el desarrollo del modelo de machine learning.

Preparar y procesar los datos históricos para el entrenamiento del modelo, aplicando técnicas de limpieza, transformación y selección de variables que garantice su calidad y relevancia.

Evaluar el rendimiento del modelo de machine learning mediante métricas de precisión, recall y F1-score, asegurando su robustez y confiabilidad en la medición de la productividad.

Diseñar estrategias de optimización basadas en los resultados del modelo, orientadas a mejorar la productividad y el desempeño del canal de ventas presencial.

Marco Teórico

Antecedentes

La medición de la productividad en equipos de ventas ha cobrado una importancia creciente en diversos sectores económicos, en especial en las telecomunicaciones debido a la alta competencia del mercado y la necesidad de optimizar resultados enfrentándose a un reto constante para mejorar su desempeño comercial basados en estrategias con datos.

La cantidad de problemas que requieren un análisis se incrementa día a día y un número significativo de ellos, son problemas de clasificación con categorías que no pueden ser definidas precisamente y las técnicas de clasificación basadas en la lógica Boole (Verdadero, Falso) presentan limitaciones (Soto & Jiménez, 2011).

De acuerdo con (Ramkumar et al., 2023) el aprendizaje automático (Machine Learning) se ha convertido en un punto clave para el análisis de datos, esto implica la utilización de diversos tipos de algoritmos que son útiles para diferentes tipos de industrias, obteniendo resultados en el aumento de sus ventas, mejora de la productividad de sus empleados y por último una muy buena rentabilidad.

Por ejemplo el alto consumo de energía en la actualidad y las emisiones de carbono han venido en constante crecimiento, lo que ha implicado en la investigación del uso de algoritmos de aprendizaje automático (Machine Learning) para predecir el consumo de energía a largo plazo en el sector de la construcción, obteniendo bases sólidas basadas en datos para que los ingenieros puedan fundamentar sus decisiones sobre sus diseños energéticos en edificios (Hussien et al., 2025).

También en el sector salud usan el aprendizaje automático (Machine Learning) para predecir a tiempo enfermedades, analizando patrones y comportamiento de los datos y

resultados de exámenes médicos e historias clínicas, por ejemplo la exposición a ftalatos y su posible vínculo con el cáncer de mama, estos se encuentran en productos de uso frecuente en mujeres, para este caso el uso del algoritmo de regresión logística multivariante, arrojó una precisión del 87,1% con lo que se concluye que estudió reveló que la exposición a ftalatos específicos aumenta el riesgo de cáncer de mama (Liu et al., 2025).

En el sector de las telecomunicaciones también es importante el uso del aprendizaje automático (Machine Learning) para resolver diferentes problemas que enfrentan las empresas a diario, con la evolución de la tecnología y la fuerte competencia se hace necesario la implementación de los algoritmos para resolver diferentes tipos de problemas. Uno de los problemas importantes que enfrentan las empresas de telecomunicaciones es la pérdida de clientes ya sea por mal servicio o una mejor oferta en la competencia y por eso es importante retenerlos ya que es fundamental para la rentabilidad de la empresa, el aprendizaje automático (Machine Learning) tiene el potencial para identificar patrones en grandes conjuntos de datos, lo que permite a las empresas anticipar la pérdida de clientes con antelación (Asif et al., 2025). Esto permite poder predecir que clientes se pueden ir de la empresa y tomar decisiones estratégicas antes de que eso suceda, estos modelos necesitan de grandes cantidades de datos, así como equipos robustos que puedan procesar toda esta información. El resultado del modelo ofrece un rendimiento robusto y una gran capacidad de interpretación, lo que destaca su potencial para apoyar los esfuerzos de retención de clientes en el TCI. (Asif et al., 2025).

En China en la industria de las telecomunicaciones durante el año 2011 y 2015 los resultados muestran que la productividad mejoró en un 22,9% anual debido al progreso tecnológico, pero también los países con niveles de desarrollo bajo de las telecomunicaciones

tienen una mayor probabilidad de impulsar su crecimiento y productividad mediante la convergencia tecnológica. (Lin et al., 2019).

Por otra parte la capacitación y el desarrollo en el desempeño de los empleados mejora la productividad de la empresa, también permite que una empresa continúe su trabajo a pesar de la rotación de personal, un estudio realizado en el sector de las telecomunicaciones en Qatar a diferentes empresas, concluyeron que el análisis muestra la solidez del vínculo entre las variables, lo que significa que las variables predictoras del modelo (Capacitación y Desarrollo) presentan una relación más estrecha con el rendimiento del empleado, lo que lleva a aceptar la hipótesis alternativa: "La capacitación y el desarrollo afectan al rendimiento del empleado".(Mekdad et al., 2024).

También se usan modelos de aprendizaje automático (Machine Learning) para analizar las quejas de los usuarios de las empresas de telecomunicaciones usando los datos generados en las interacciones con los clientes, el estudio destaca la importancia de la preparación de los datos en el entrenamiento del modelo y el resultado es información valiosa para creación de estrategias enfocadas en el servicio al cliente (Jaber et al., 2023).

Los equipos comerciales de las empresas son los más importantes y tienen un papel fundamental en las estrategias de la empresa, ya que la fuente de ingresos, su desarrollo y crecimiento están relacionados con los intereses de la empresa, por esta razón es importante medir la productividad de acuerdo a los resultados en ventas, estos deben ser medidos de acuerdo a las metas comerciales fijadas por la empresa, sin embargo este enfoque es poco común y muchas empresas manejan modelos de comisión por ventas, entre más ventas más dinero ganan lo que hace atractivo el cargo, por esta razón es que la mayoría de estudios utilizan resultados enfocados a procesos y desarrollo del empleado, (Cai et al., 2022).

Marco Conceptual

La productividad se define como la relación de resultados comerciales obtenidos por la fuerza comercial y los recursos utilizados, el canal de ventas presencial es un conjunto de interacciones entre los vendedores y los clientes. En este contexto se utilizará el aprendizaje automático (Machine Learning) para aprender a partir de los datos históricos con el fin de predecir la productividad.

El machine learning es un subcampo de la inteligencia artificial que ayuda a los computadores a aprender y actuar como seres humanos con la ayuda de algoritmos y datos. Un algoritmo de datos aprende diferentes propiedades de los datos e infiere las propiedades de los datos que se pueden presentar en el futuro (Jiménez, 2021).

Teniendo en cuenta lo anterior los datos son una parte fundamental del aprendizaje automático (Machine learning), son conjuntos de datos almacenados en diferentes sistemas de gestión de base de datos (SGBD), estos pueden ser estructurados o no estructurados y permite a los usuarios definir, crear y mantener la base de datos, además de proporcionar acceso controlado a la misma (*Bases de Datos*, s. f.)

Hoy en día es más frecuente la generación masiva de datos de diferentes tipos como económicos, comerciales, personales, financieros, imágenes entre otros lo que implica una cantidad bastante grande de información y su peso se puede medir en Terabytes y si hablamos de grandes empresas pueden ser Petabits de información y se hace necesario la minería de datos, es una técnica asistida para procesar y explorar grandes conjuntos de datos, de acuerdo con Amazon los sectores verticales de alta competencia, como las telecomunicaciones, los medios y la tecnología utilizan la minería de datos para mejorar el servicio de atención al cliente mediante la búsqueda de patrones en el comportamiento de estos. Por ejemplo, una empresa podría analizar

los patrones de uso del ancho de banda y proporcionar actualizaciones o recomendaciones de servicio personalizadas (*¿Qué es la minería de datos?*, s. f.).

El aprendizaje automático (Machine Learning) está dedicado al desarrollo de algoritmos que pueden extraer patrones de diferentes tipos de datos, por ejemplo algoritmos de clasificación o de predicción, estos procesos no son automatizables ya que se debe realizar una tarea específica con los datos como es la validación de la calidad de los datos, estructurarlos, en algunos casos deben ser completados porque los registros llegan vacíos, hacer transformaciones, normalizar datos, aplicar análisis estadísticos para no tener en cuenta las variables que tienen una correlación bastante fuerte y a partir de estos resultados se puede hacer la elección del algoritmo más adecuado para el caso (Bobadilla, 2021) los siguiente son los tipos de modelos:

Figura 1

Tipos de Modelos

- ▼ Aprendizaje supervisado
 - regresión
 - clasificación
- ▼ Aprendizaje no supervisado
 - clustering (agrupamiento)
 - reducción de dimensiones
- ▼ Aprendizaje semi-supervisado
- ▼ Aprendizaje por refuerzo

Nota. Tomado de Machine Learning y Deep Learning: Usando Python, Scikit y Keras

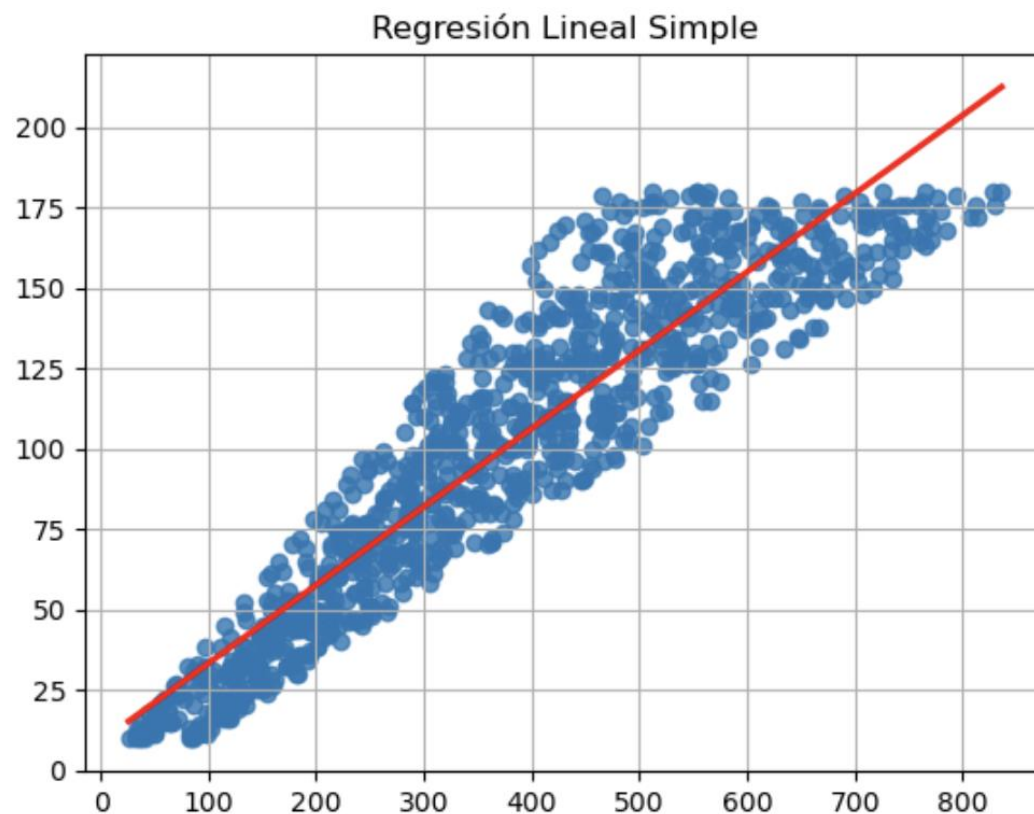
Cuando se usa los modelos de aprendizaje supervisado el conjunto de datos con el que se va a trabajar tiene asociada una etiqueta, por ejemplo un conjunto de datos de una liga de futbol, la etiqueta seria si un equipo fue campeón o no de los torneos anteriores, esto con el fin de poder

entrenar el modelo y poder predecir la etiqueta correspondiente al torneo actual, es decir cuál de los equipos sería el campeón de la liga, también se pueden usar estos modelos para predecir el precio de una vivienda, sistemas de recomendación entre otros.

Uno de los algoritmos más usados es la regresión lineal que está dirigida a describir como es la relación entre dos variables (X,Y) y también puede hacer predicciones sobre los valores de la variable (Y), a partir de los datos de (X). Si tenemos dos variables hablamos de una regresión lineal simple y si hay más de dos variables es una regresión lineal múltiple (Laguna, s. f.) en la siguiente figura se puede ver el resultado de un modelo de regresión lineal simple donde se puede evidenciar la relación de dos variables

Figura 2

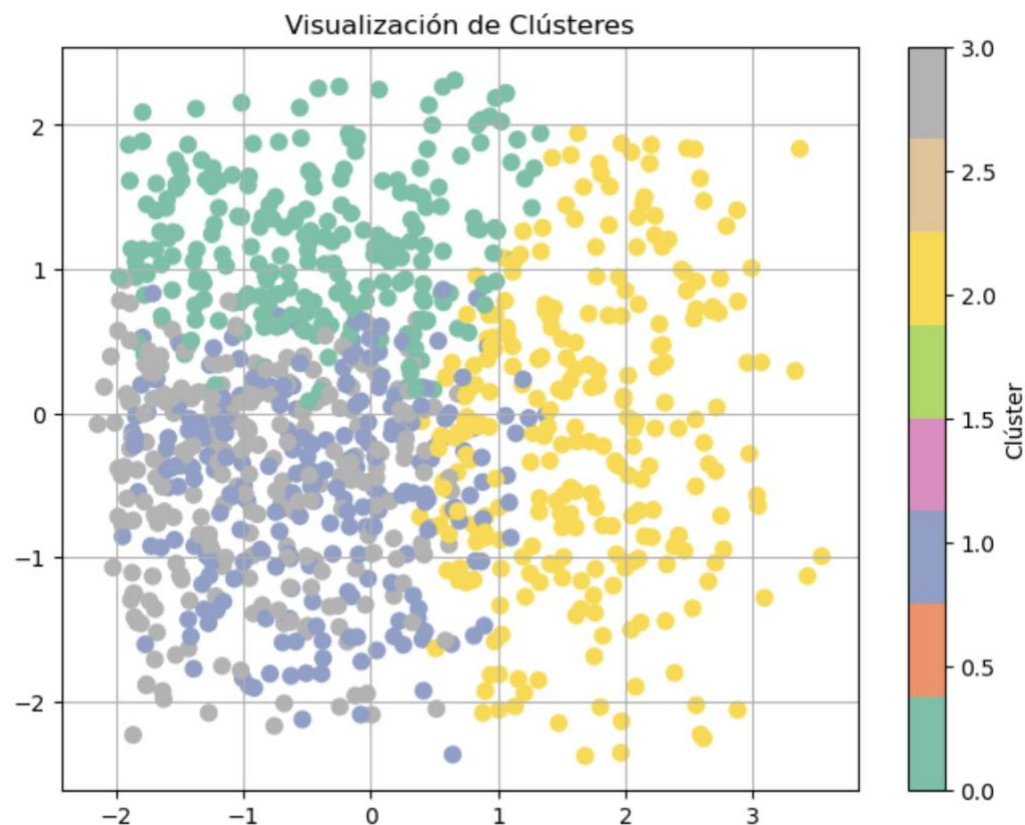
Regresión Lineal Simple



En cuanto al aprendizaje no supervisado los diferentes algoritmos utilizan datos no etiquetados, los más conocidos son los clustering (Agrupamiento) y su objetivo es agrupar muestras, por ejemplo se tiene la data de los empleados de una empresa con su información socio demográfica el clúster va a realizar agrupamientos de acuerdo a los datos suministrados, puede hacer agrupaciones por género, edad, profesión entre muchas otras, también hay diferentes clústeres que proporcionan elementos virtualmente representativos que son los “centroides” cuya área de influencia determina a qué clúster pertenece cada una de las muestras, como decía hay varios tipos de clústeres y se usan de acuerdo a los datos y necesidad que se tenga (Bobadilla, 2021) en la siguiente figura se puede ver un ejemplo de un gráfico de clúster.

Figura 3

Clúster K-Means



“El aprendizaje semi-supervisado trata con conjuntos de datos en los que una porción de los datos está etiquetada y el resto no. Normalmente, la cantidad de muestras etiquetadas es mucho más pequeña que las no etiquetadas. La mayoría de los algoritmos de aprendizaje semi-supervisado son una mezcla de métodos supervisados y no supervisados” (Bobadilla, 2021)

Los árboles de decisión son un modelo predictivo y de clasificación, estos son muy similares a sistemas de predicción basado en reglas, también los modelos clasifican casos en grupos de una variable dependiente, es uno de los modelos más populares y usados en la actualidad por su inteligibilidad y simplicidad, lo árboles se representan gráficamente por nodos, hojas y ramas, donde el nodo principal es el atributo o variable importante que inicia el proceso de clasificación, también existen nodos internos que corresponden a un evento aleatorio o pregunta acerca de los atributos, la respuesta a estas preguntas se representan por un nodo hijo y es la probabilidad de que un evento suceda, las ramas que cada nodo tiene son etiquetadas con posibles valores de los atributos y finalmente están los nodos finales que corresponden a una decisión (Ver Figura 4). (Martínez et al., 2009).

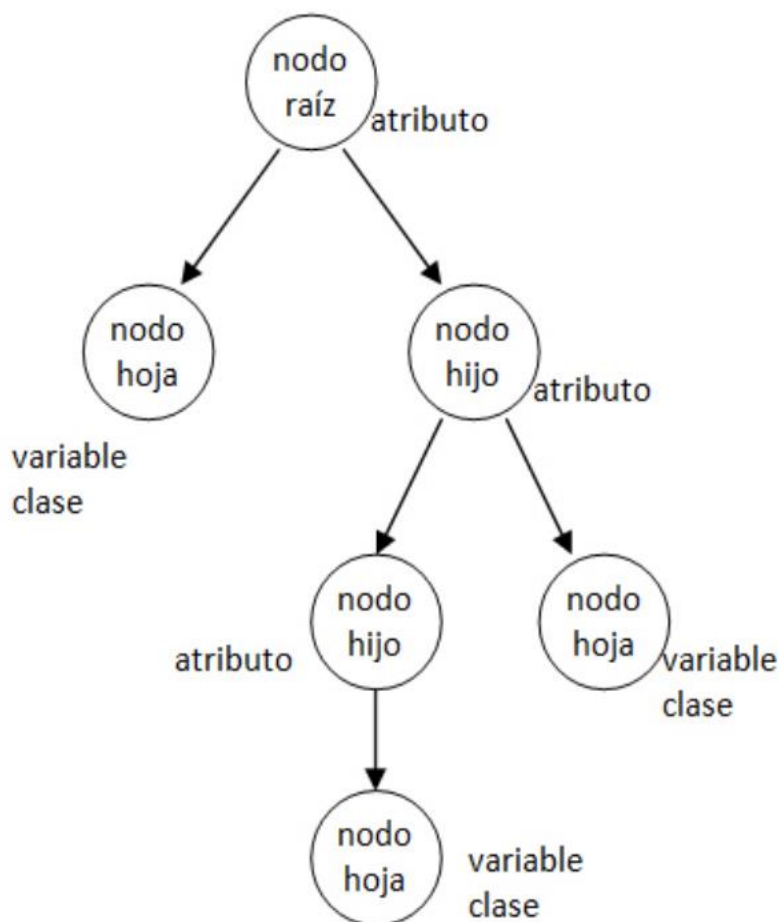
Las ventajas que tienen los árboles de decisión son su fácil interpretación de la decisión adoptada, explica el comportamiento respecto a una determinada decisión y también reduce el número de variables independientes, otra de las ventajas es que es capaz de manejar datos categóricos y numéricos a diferencia de otros modelos que solo usan datos números y se debe hacer la conversión de datos categóricos a numéricos 0 y 1, en cuanto a la cantidad de datos funciona bastante bien con conjuntos de datos grandes.

En este contexto, se utilizara el modelo de predicción de árbol de decisión partiendo de datos históricos del canal de ventas presencial de una empresa de telecomunicaciones, con el fin de aprender patrones y predecir la productividad de los asesores del canal, el modelo incluirá

variables clave que son cuantitativas y cualitativas, permitiendo estructurar un modelo de algoritmo supervisado, lo que va a permitir la toma de decisiones estratégicas, en la siguiente figura se puede ver la estructura del árbol.

Figura 4

Estructura de un Árbol de Decisión



Nota. Tomado de Árboles de decisión como herramienta en el diagnóstico médico

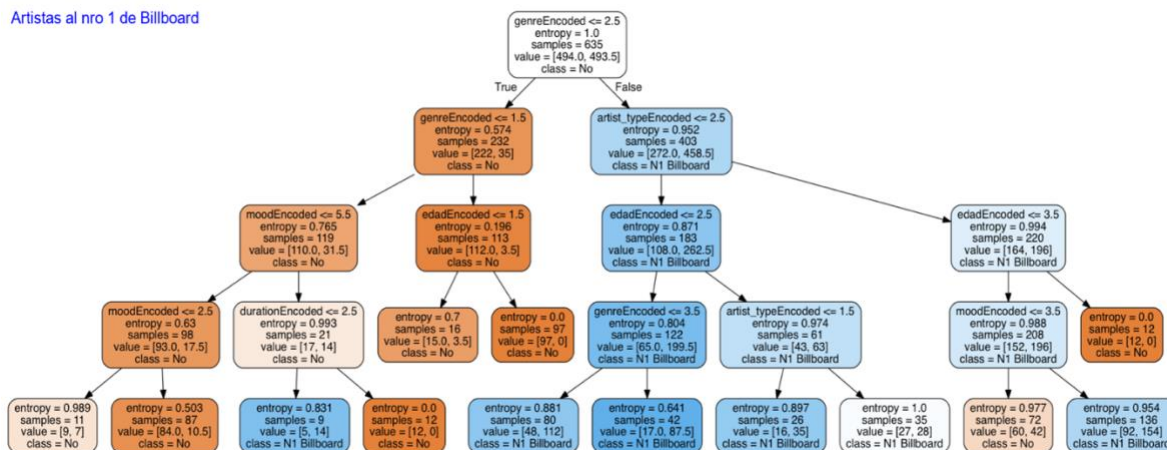
Los algoritmos de árbol se pueden trabajar en dos etapas, en la primera etapa se entrena el conjunto de datos, estos se dividen en entrenamiento y prueba y se pueden aplicar técnicas para el balanceo de datos, en caso de que se necesite, como también si hay desequilibrio en los

datos, esto con el fin de no tener un modelo sesgado, estas técnicas solo deben aplicarse al conjunto de datos de entrenamiento, no se deben aplicar sobre el conjunto de datos de prueba o test, en caso de aplicar alguna técnica es importante hacer una comparación del antes y después de aplicarlas, con el modelo entrenado se crea un árbol y se realiza la evaluación del modelo, esto permite identificar si es necesario realizar algún tipo de ajuste sobre el entrenamiento. En la segunda etapa se crea el árbol de decisión se crea el árbol y se ajustan los parámetros de control, en estos parámetros se puede ajustar por ejemplo la profundidad máxima del árbol, cuantos registros mínimos puede tener una hoja entre otros, después de ese proceso ya se puede visualizar el resultado de un árbol, en la siguiente figura se puede ver un ejemplo del resultado de un modelo de árbol.

Figura 5

Árbol de Decisión Artistas al 1 de Billboard

Artistas al nro 1 de Billboard



Nota. Tomado de Aprende Machine Learning

“Para obtener el árbol óptimo y valorar cada subdivisión entre todos los árboles posibles y conseguir el nodo raíz y los subsiguientes, el algoritmo deberá medir de alguna manera las

predicciones logradas y valorarlas para comparar de entre todas y obtener la mejor.” (*Crea un Arbol de Decisión en Python | Aprende Machine Learning*, s. f.)

La implementación del modelo de árbol de decisión se va a desarrollar en Python, un lenguaje de programación sencillo, liviano e ideal para aprender, esta herramienta cuenta con librerías especializadas en la ciencia de datos, lo que hace que sea robusto y eficiente a la hora de procesar datos y crear modelos de aprendizaje automático (Machine Learning), es una herramienta de código libre que no solo permite la creación de algoritmos sino que también permite la limpieza y transformación de datos, las librerías usadas con más frecuencia son Pandas, Numpy, scikit-learn, Matplotlib y seaborn, estas dos últimas son especializadas en visualizaciones.

La elección de Python en el proyecto se justifica por su capacidad de trabajar con grandes cantidades de datos y realizar operaciones matemáticas complejas, el flujo de trabajo es bastante bueno y permite hacer ajustes de manera muy sencilla, su entorno es bastante amigable y los recursos que consume del computador no son muchos, claramente depende del algoritmo que se procese y la cantidad de datos, dependiendo de la complejidad del algoritmo se necesitan buenos recursos a nivel de procesador, memoria RAM y GPU, es recomendable procesar estos algoritmos en servidores o en la nube, en el computador personal se puede diseñar el modelo y ajustarlo para su procesamiento.

Entrando un poco más en contexto de Python fue introducido en 1991, desarrollado por el programador holandés Guido van Rossum, mientras él trabaja en Google se le permitió la mitad de sus horas laborales trabajar en el desarrollo de Python, entre en el 2005 a 2012 ya se había convertido en uno de los lenguajes más populares, los objetivos de Python eran ser fácil de aprender, código legible, intuitivo y algo muy importante que fuera de código abierto.

Finalmente para entrenar el modelo de árbol de decisión, se seleccionaron variables clave que representan la operación del canal de ventas presencial, estas variables se definieron en base al conocimiento del negocio y la disponibilidad de datos históricos, la variable objetivo es la **productividad**, categorizada en niveles que permiten interpretar y clasificar el rendimiento comercial: *aceptable, retroalimentación, improductivo y productivo*, la descripción detalla de las variables utilizadas se presentan a continuación:

Tabla 1

Descripción de las Variables

Variable	Tipo de Dato	Descripción	Rol
Contrato	Categórica	Tipo de contrato	Predictora
Pago	Numérica	Variable o fijo	Predictora
Antigüedad	Categórica	Tiempo en la empresa	Predictora
Ventas hogar	Numérica	Servicios fijos	Predictora
Ventas pospago	Numérica	Servicios móviles	Predictora
Ventas TyT	Numérica	Terminales y tecnología	Predictora
Retención	Numérica	Fidelización de clientes	Predictora
NPS	Numérica	Satisfacción del cliente	Predictora
TPA	Numérica	Tiempo prom de atención	Predictora
Turnos	Numérica	Clientes atendidos	Predictora
Generación	Categórica	Generación (Gen X)	Predictora
Categoría	Categórica	Tipo de oficina	Predictora
Productividad	Categórica	Objetivo a predecir	Variable Objetivo

Nota: Descripción de las variables usadas en el modelo, elaboración propia

Las variables seleccionadas para el proyecto permite estructurar un modelo robusto, donde los datos históricos son fundamentales y alimentan el algoritmo de árbol de decisión, que no solo va a predecir la productividad sino que también va a proporcionar información valiosa

para diseñar estrategias y la toma de decisiones estratégicas en tiempo real, teniendo esto en cuenta se refuerza el objetivo del proyecto: mejorar la eficiencia comercial mediante el uso de técnicas avanzadas de analítica y aprendizaje automático (Machine Learning).

Marco Legal

Este proyecto se acoge a lo dispuesto en la ley 1581 de 2012 de Colombia, por medio de la cual se dictan disposiciones generales para la protección de datos personales. En particular, se siguen los principios de legalidad, finalidad, seguridad y confidencialidad en el tratamiento de la información recolectada de los asesores del canal de ventas presencial. Según el artículo 10, literal b), no se requiere autorización previa para el tratamiento de datos con fines estadísticos o científicos, como es el caso de este modelo de productividad, siempre que se garantice el anonimato de los individuos y el uso interno de los resultados. Así, el desarrollo del proyecto se realiza bajo un marco legal que protege los derechos de los titulares de los datos y asegura un uso responsable de la información (Congreso de Colombia, Ley 1581, 2012)

La ley 1581 de 2012 establece:

Artículo 4°. Principios para el Tratamiento de datos personales.

En el desarrollo, interpretación y aplicación de la presente ley, se aplicarán, de manera armónica e integral, los siguientes principios:

a) Principio de legalidad en materia de Tratamiento de datos: El Tratamiento a que se refiere la presente ley es una actividad reglada que debe

sujetarse a lo establecido en ella y en las demás disposiciones que la desarrollen;

b) Principio de finalidad: El Tratamiento debe obedecer a una finalidad legítima de acuerdo con la Constitución y la Ley, la cual debe ser

informada al Titular;

c) Principio de libertad: El Tratamiento sólo puede ejercerse con el consentimiento, previo, expreso e informado del Titular. Los datos personales

no podrán ser obtenidos o divulgados sin previa autorización, o en ausencia de mandato legal o judicial que releve el consentimiento;

g) Principio de seguridad: La información sujeta a Tratamiento por el Responsable del Tratamiento o Encargado del Tratamiento a que se refiere

la presente ley, se deberá manejar con las medidas técnicas, humanas y administrativas que sean necesarias para otorgar seguridad a los

registros evitando su adulteración, pérdida, consulta, uso o acceso no autorizado o fraudulento;

h) Principio de confidencialidad: Todas las personas que intervengan en el Tratamiento de datos personales que no tengan la naturaleza de

públicos están obligadas a garantizar la reserva de la información, inclusive después de finalizada su relación con alguna de las labores que

comprende el Tratamiento, pudiendo sólo realizar suministro o comunicación de datos personales cuando ello corresponda al desarrollo de las

actividades autorizadas en la presente ley y en los términos de la misma. (Congreso de Colombia, Ley 1581, 2012)

Desarrollo

Fase 1 Definición de las Variables Clave

En el contexto de las ventas la productividad se define como un índice que relaciona los resultados obtenidos como el número de ventas, ingresos generados y los recursos usados para producirlos como el tiempo invertido o las herramientas usadas, entre otros, esto se refiere a la eficiencia con la que se utilizan estos recursos para el cumplimiento de objetivos comerciales, permitiendo identificar áreas de mejora y la toma de decisiones estratégicas (*Productividad y competitividad*, s. f.).

Para poder definir las variables que se van a usar en el modelo se hizo un análisis 360 en la operación, se validaron todos los indicadores comerciales como de servicio que actualmente se miden en todo el canal de ventas presencial, de esta manera se empezó a consolidar las posibles variables, también se validaron los reportes comerciales vigentes y una validación a los reportes históricos.

Teniendo en cuenta lo anterior se programaron sesiones con el equipo comercial, el equipo de datos y el equipo operativo con el fin de validar conceptos y conocer las variables clave para cada uno de estos equipos, se recibieron sugerencias y recomendaciones de acuerdo con su experiencia y conocimiento del canal de ventas presencial.

Con la información recibida y teniendo una vista más clara y alienada de la operación, se identificaron variables que miden el desempeño integral de un asesor de ventas del canal de ventas presencial, cada variable se seleccionó basándose en su relevancia operativa y en la disponibilidad histórica de la información para que el modelo sea robusto, en la siguiente tabla se describe cada variable, su tipo de dato y la justificación de su elección para el modelo de aprendizaje automático (Machine Learning).

Tabla 2*Justificación de las Variables*

Variable	Tipo de dato	Justificación
Contrato	Cualitativo	Diferentes tipos de contrato se puede asociar a niveles de compromiso y estabilidad
Pago	Cuantitativo	Influye en la motivación y comportamiento de ventas
Antigüedad	Cualitativo	Experiencia y conocimiento, factor que puede elevar la productividad
Ventas hogar	Cuantitativo	Mide el desempeño en las ventas de servicios fijos
Ventas pospago	Cuantitativo	Mide el desempeño en las ventas de servicios móviles
Ventas TyT	Cuantitativo	Mide el desempeño en las ventas de equipos terminales y tecnología
Retención	Cuantitativo	Factor crítico en la rentabilidad
NPS	Cuantitativo	Calidad del servicio y la probabilidad de ser recomendados, ventas futuras
TPA	Cuantitativo	Eficiencia en la atención de clientes
Turnos	Cuantitativo	Mide el desempeño y carga laboral
Generación	Cualitativo	Influyen en los estilos de ventas y adopción de conocimiento
Categoría	Cualitativo	Categoría de cada oficina presencial
Productividad	Cualitativo	Clasifica el desempeño y entrena el modelo para predecir la clasificación de la productividad.

Nota: Justificación de la elección de las variables

Con la definición de las variables clave para el modelo respaldadas por la justificación en la tabla anterior, ya se cuenta con la base sólida para el desarrollo del modelo de aprendizaje automático (Machine Learning), el conjunto de datos con las variables seleccionadas ya está documentado y listo para continuar con las siguientes fases.

Fase 2 Recolección y Preparación de Datos

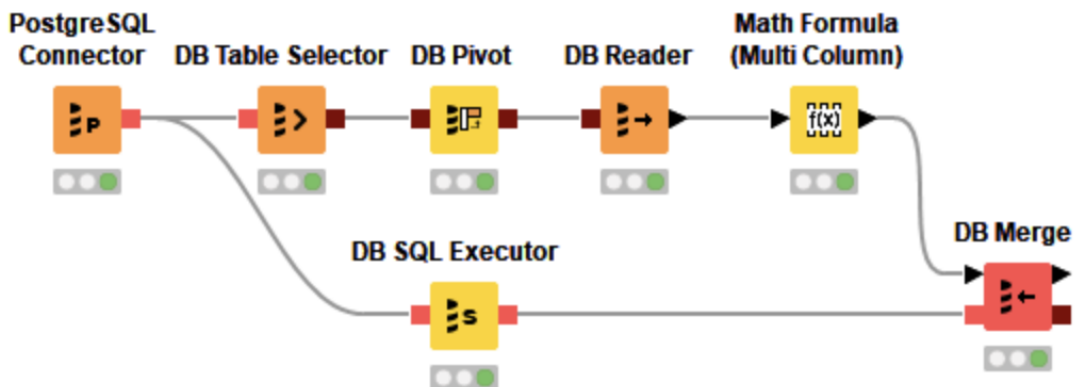
Con las variables ya definidas para el modelo, en esta segunda fase se centra en reunir los datos históricos de cada variable para transfórmalos y normalizarlos si es el caso, esta fase es la más importante del proyecto ya que un dataset bien estructurado permite tener un modelo robusto y no sesgado.

Para la recolección de los datos históricos fue necesario realizar la conexión a los diferentes sistemas de la empresa y también hacer la conexión con el Data Warehouse (DWH). Con las conexiones habilitadas a las fuentes de datos se creó una ETL (Extracción, Transformación y Carga) para poder extraer la información y realizar las diferentes transformaciones y finalmente realizar el cargue en SQL Server para tener los datos estructurados y consolidar toda la información en una sola tabla.

Para la creación de la ETL (Extracción, Transformación y Carga) se usó un software de código abierto KNIME que integra la minería de datos, ciencia de datos, análisis de datos, modelado y visualización, KNIME es una herramienta muy sencilla de usar, además de ser gráfica integra nodos con diferentes tipos de acciones que se pueden ejecutar sobre los datos y flechas para conectar los nodos representando un flujo de datos, tiene la posibilidad de conectarse con Python y R para ejecutar scripts, lo que lo hace una muy buena herramienta no solo para extraer y transformar información sino también para la automatización de procesos que son bastante operativos para muchas áreas en diferentes empresas.

Figura 6

Ejemplo de Flujo de Trabajo con Nodos de Bases de Datos



Nota. Tomado de Documentación Knime

En la figura 6 se puede ver un ejemplo de cómo es la estructura de un flujo de KNIME con las bases de datos y los diferentes nodos que están disponibles para usar, en este caso están ejecutando diferentes tareas conectándose por medio de JDBC a una base de datos en Postgre. En la siguiente figura se puede ver el diagrama de una ETL.

Figura 7

Diagrama ETL



Para consolidar la información en una única tabla, se implementaron seis procesos ETL encargados de extraer, transformar y cargar los datos. Partiendo de las bases sociodemográficas de los asesores de ventas, se utilizó la cédula como llave primaria para relacionarla con las demás bases, todas las tablas cuentan con sus campos poblados, por lo que no es necesario validación de nulos, garantizando así la calidad y consistencia de la información.

Dentro del proceso de ETL, se aplicó la técnica de codificación de variables cualitativas mediante Label Encoding usando la librería scikit-learn para Python, a dos variables categóricas. Este proceso consiste en asignar un valor numérico a cada categoría de una variable. Por ejemplo, si una variable tiene dos categorías como "Sí" y "No", se les asignaría el valor numérico 0 a "Sí" y 1 a "No".

Label Encoding es una técnica bastante eficiente, ya que transforma rápidamente variables categóricas en números, lo que facilita su procesamiento por algoritmos de aprendizaje automático (Machine Learning). Es bastante útil por su simplicidad y bajo coste computacional, lo que permite trabajar con grandes volúmenes de datos de manera eficaz obteniendo muy buenos resultados en el modelo.

En conclusión, ya se han construido los flujos de datos para las ETLs, consolidando la información. Además, consolidaremos los datos de los últimos tres meses de cada uno de los asesores de ventas, con el fin de evitar sesgos en la información y minimizar las posibles variaciones debido a la estacionalidad del canal. La consolidación de la información de las variables de ventas y servicio se realizó promediando la información de los últimos tres meses, así como las metas comerciales asignadas. Esto con el fin de evaluar el cumplimiento de cada indicador (Ventas / Metas Comerciales) y clasificarlos en la variable **Productiva**, que es la

variable objetivo. De acuerdo con estos resultados los asesores pueden ser clasificados en las siguientes categorías: *aceptable*, *retroalimentación*, *improductivo* y *productivo*.

Una vez consolidada la información en una tabla de SQL Server, se procedió a exportarla en un archivo de texto delimitado por comas (CSV) para su análisis en Python, en esta etapa se realizó un análisis exploratorio de datos (EDA) con el fin de verificar la consistencia de la información consolidada.

Se analizaron las variables numéricas para evaluar su comportamiento y detectar valores atípicos o inconsistencia con los datos, para esto se generaron visualizaciones como diagramas de cajas (Boxplots) e histogramas, en caso de encontrar valores que afecten el modelo es necesario aplicar técnicas de normalización o estandarización para no afectar el algoritmo.

En este caso no aplicamos ninguna técnica de normalización o estandarización en los datos, ya que el dataset no presenta sesgo ni datos atípicos.

Después se realiza la división del conjunto de datos para entrenar el modelo aplicando la técnica Train Test Split de la librería scikit-learn para Python, esta técnica divide el conjunto de datos, en dos partes, una parte para entrenar el modelo y la otra parte para el test del modelo, para este caso se dividió el conjunto de datos en 80% entrenamiento y 20% test, después aplicamos la técnica de SOMTE que se utiliza para balancear el conjunto de datos, esta técnica se debe aplicar sobre el conjunto de datos entrenados, se usa para equilibrar el conjunto de datos debido a una categoría minoritaria de la variable objetivo, esto quiere decir que en el conjunto de entrenamiento se encontraron pocas muestras de la categoría “*Improductivo*” y “*Retroalimentación*” entonces sería muy sencillo para el modelo predecir las categorías “*Productivo*” y “*Aceptable*” la técnica genera ejemplos sintéticos (Falsos) tomando una muestra de la clase minoritaria y encontrando sus vecinos cercanos y selecciona uno al azar y genera un

punto adicional entre la instancia original y el vecino. Esta técnica mejora la capacidad del modelo para aprender características de la clase minoritaria, también reduce el sobreajuste que pasa cuando se duplican las clases minoritarias y por último mejora las métricas de recall o F1-Score.

Finalmente se hace la validación de la distribución de las clases antes y después del balanceo para validar que la aplicación de la técnica SMOTE se hizo de manera correcta, a continuación, vemos los resultados en la siguiente tabla:

Tabla 3

Balanceo de Datos

Balanceo	Clase 0	Clase 1	Clase 2	Clase 3
Antes	773	53	160	454
Después	773	773	773	773

Nota: Distribución del balanceo de datos por clase, antes y después de aplicar SMOTE

Finalizando la recolección, consolidación y preprocesamiento de los datos, se cuenta con un dataset estructurado, limpio, entrenado y balanceado que respeta la definición de las variables establecidas en la fase 1, el dataset no contiene valores nulos ni inconsistencias. El balanceo con la técnica SMOTE garantiza que no se va a sesgar las predicciones a una clase mayorista. Asimismo, el 20% de los datos se preserva para la evaluación objetiva del desempeño.

Fase 3 Desarrollo y Evaluación del Modelo

Partiendo de un dataset previamente balanceado, nos enfocaremos en la construcción, calibración y validación del árbol de decisión diseñado para predecir la productividad de los asesores del canal presencial, el objetivo del modelo es alcanzar métricas de precisión y robustez adecuadas para su aplicación práctica.

Inicialmente creamos el árbol de decisión para la clasificación (DecisionTreeClassifier) y configuramos algunos hiperparámetros específicos que nos permiten controlar el crecimiento del árbol y su comportamiento, para este proyecto configuramos 5 hiperparámetros para asegurar que el modelo no tenga sobreajuste, limitando la profundidad del árbol y el tamaño mínimo de las hojas, se usó la entropía (Entropy) para decidir cómo dividir los nodos y una profundidad máxima, en la siguiente tabla está la explicación de los hiperparámetros:

Tabla 4

Explicación de los Hiperparámetros

Parámetro	Selección	Explicación
Criterion	Entropy	Función para medir la calidad de una división
min_samples_split	20	Mínimo de muestras para dividir un nodo
min_samples_leaf	5	Mínimo de muestras en una hoja del árbol
max_depth	4	Establece la profundidad máxima del árbol
random_state	42	Semilla aleatoria para asegurar resultados reproducibles

Nota: Parámetros aplicados en el modelo, elaboración propia

Una vez ajustados los hiperparámetros, se logra reducir el riesgo de sobreajuste y se controla la profundidad del árbol, lo que contribuye a tener un modelo con mejor desempeño. A continuación, se procede a entrenar el modelo con el conjunto de datos de entrenamiento y luego se realizan las predicciones sobre el conjunto de test o validación, con el fin de evaluar su rendimiento.

Para la evaluación de rendimiento del modelo, se calcularon las métricas de precisión, recall y f1-score, estas métricas miden la efectividad del modelo al clasificar correctamente las instancias, a continuación, una explicación de las métricas:

1. **Precisión:** Indica el porcentaje de predicciones positivas que son correctas, mide la exactitud del modelo.
2. **Recall:** Evalúa la capacidad del modelo para detectar todos los casos reales.
3. **F1-Score:** Es la media entre precisión y recall, proporciona equilibrio entre ambas métricas.

A continuación, el resultado de las métricas de evaluación para el modelo de productividad del canal de ventas presencial:

Tabla 5

Métricas de Evaluación del Modelo

Variable Objetivo	Precisión	Recall	F1-Score
Aceptable	0.81	0.75	0.70
Retroalimentación	0.85	0.79	0.81
Improductivo	0.82	0.82	0.73
Productivo	0.80	0.81	0.72
Macro AVG	0.82	0.79	0.74

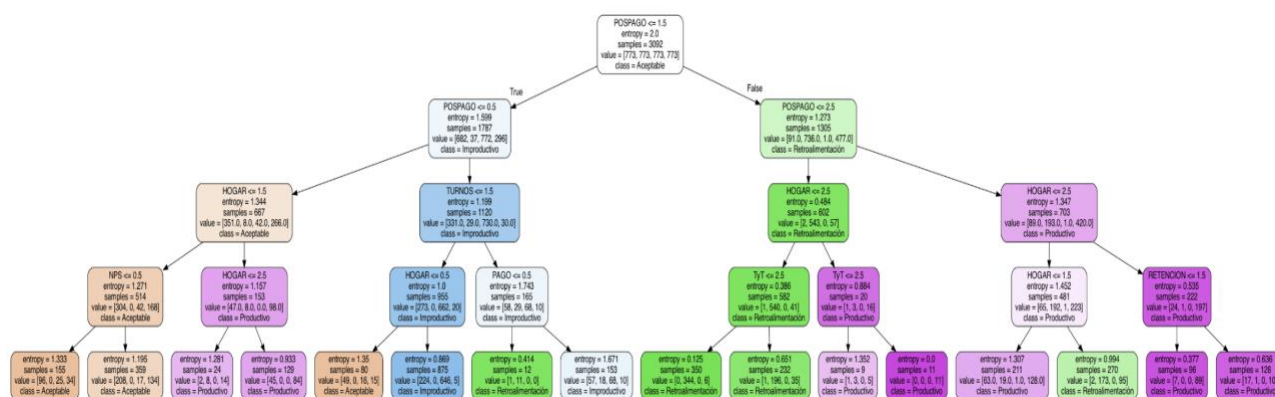
Nota: Métricas usadas para evaluar el modelo, elaboración propia

De acuerdo con los resultados de la evaluación del modelo se obtuvo una macro precisión de 0.82 y un macro recall de 0.79, esto nos indica que el modelo rinde de manera homogénea en todas las clases. Además, el F1-score macro fue de 0.74, lo que refuerza la conclusión de que el modelo presenta un buen desempeño general, incluso en un contexto multiclase.

Estos resultados cumplen con niveles de rendimiento “buenos” aportando confianza para su uso en la toma de decisiones, incluso logra un equilibrio entre precisión y recall incluyendo la categoría “Aceptable”, tradicionalmente más compleja de distinguir.

Figura 8

Resultado del Árbol de Decisión



Finalmente, en esta etapa se construyó y optimizó un árbol de decisión que es capaz de predecir la productividad de los asesores de ventas del canal presencial en cuatro categorías “Retroalimentación”, “Improductivo”, “Aceptable” y “Productivo” con un macro F1-Score de 0.74 que garantiza un desempeño equilibrado. Con un modelo robusto y bien calibrado, contamos ahora con una herramienta predictiva validada que refleja con fidelidad los patrones de desempeño en el canal presencial. Este resultado no solo permite anticipar el rendimiento de los asesores, sino que también sienta las bases para futuras estrategias de segmentación y mejora continua en la gestión del talento comercial.

Fase 4 Estrategias de Optimización

Con el modelo de árbol de decisión validado y optimizado en la fase anterior, ahora contamos con una herramienta predictiva que es capaz de identificar con un buen grado de certeza a los asesores con distinta categorización de productividad. Esta última fase se enfoca en poder transformar esas predicciones en decisiones estratégicas y planes de acción concretos, partiendo de los patrones detectados como identificando las variables que más influyen en la improductividad y que afectan el rendimiento óptimo de los asesores. El objetivo no solo es

mejorar los indicadores de productividad, sino establecer un proceso continuo de retroalimentación que permita adaptar de manera dinámica estrategias durante el mes en curso y garantizar una eficiencia del canal presencial de ventas.

Teniendo en cuenta los resultados arrojados por el modelo podemos identificar la ruta principal hacia la categoría “*Productivo*” donde las ventas de la variable “pospago” altas o moderadas combinadas con métricas de “NPS”, “Retención”, “Turnos” y las características como contrato o generación empujan al árbol hacia “Productivo”.

Un equilibrio entre las ventas “Pospago” y “Hogar” donde no es necesario que las ventas sean específicamente altas en estas dos variables, sino moderadas apoyado por buenas métricas de servicio como el “NPS” y una carga laboral como “Turnos”, indicadores como “Retención” y “TPA” afinan la predicción y nos da la ruta hacia una buena productividad.

De acuerdo con la ruta hacia la productividad podemos enfocarnos en las variables que pueden estar afectando el desempeño de algunos asesores de ventas, muchas ramas del árbol conducen hacia la clasificación “Retroalimentación” cuando el “NPS” es bajo y los “Turnos” son bastante altos, lo que nos está indicando una mala atención en servicio al cliente. En cuanto a los “Aceptables” que es un número importante se clasifica así cuando las ventas de “Terminales y Tecnología” son bajas, pero sin garantía suficiente en las variables de “NPS” y “Retención”, esto nos indica que es un perfil potencial para ser productivo.

También se analizó la importancia de las variables utilizadas en la predicción, este análisis permite identificar que atributos tienen mayor peso en la productividad de los asesores del canal de ventas presencial. A continuación, los resultados obtenidos:

Tabla 6*Importancia de las Variables en el Modelo*

Variable	Importancia
Pospago	0.373
Hogar	0.278
TyT (Terminales y Tecnología)	0.140
Turnos	0.054
NPS	0.044
Pago	0.027
TPA	0.027
Retención	0.016
Generación	0.014
Categoría	0.011

Nota: Importancia de las variables del modelo. Elaboración propia

Los resultados mostraron que las variables “Pospago” y “Hogar” concentran más del 60% de la importancia del modelo y son un factor determinante en su rendimiento, la variable “TyT” aporta de forma significativa seguida por las demás variables “NPS”, “Pago”, “TPA” que si bien no tienen mucha importancia son relevantes para el modelo. Esto es clave en la interpretación del modelo ya que permite focalizar estrategias sobre los segmentos más influyentes sin dejar a un lado las menos importantes, pero con oportunidades de optimización, que podrían mejorarse y destacarse en futuras iteraciones del modelo.

De acuerdo con los resultados obtenidos, se debe hacer la implementación de un dashboard que permita hacer seguimiento a la productividad por asesor en tiempo real, donde se pueda monitorear todas las variables del modelo, este dashboard se puede implementar en

herramientas como Power BI o Tableau, permitiendo hacer seguimiento de manera muy sencilla y generando alerta sobre los indicadores que necesiten atención.

Con el dashboard implementado, se pueden identificar los grupos de asesores con mayor probabilidad de ser “Improductivos” o “Retroalimentación” con los que el supervisor o coordinador de la oficina debe empezar a trabajar para mejorar su rendimiento durante el mes en curso y realizar retroalimentación efectiva.

Se propone crear planes de formación enfocados en las variables con menos importancia con el fin de fortalecerlas y que sean importantes a futuro, estos planes de formación pueden estar enfocados en la gestión del tiempo “TPA” para aquellos asesores que tienen tiempos muy altos y un mal “NPS”. También planes de formación enfocados a la “Retención” del cliente ya que son importantes para la compañía, así como el servicio al cliente, que puede ser en el manejo de objeciones.

Otra estrategia para mejorar el desempeño de los asesores es la implementación de incentivos, que no solo motivan sino impactan en el rendimiento individual y colectivo, los incentivos bien dirigidos pueden actuar como mecanismo de mejora continua al equipo comercial. Con base en la predicción del modelo se propone estructurar un esquema de incentivos segmentado por la clasificación de la “Productividad” que podría ser de la siguiente manera:

Tabla 7

Estructura de los Incentivos

Productividad	Incentivo
Improductivo	Programa de mentoría y coaching
Retroalimentación	Compromiso de mejora y acceso a recompensas bajas
Aceptable	Recompensas en dinero menores, por mejora continua

Nota: Propuesta de los incentivos por clasificación. Elaboración propia

Por último, el ciclo de retroalimentación del modelo es bastante importante ya que se debe alimentar de los datos actuales y reentrenarlo de manera mensual para que siga realizando las predicciones de manera eficaz y con más precisión teniendo en cuenta la aplicación de las estrategias y la mejora de los asesores de ventas.

Los resultados esperados después de aplicar las estrategias son el incremento medible de F1-score en las categorías “Aceptable” y “Retroalimentación”, un mayor número de asesores clasificados como “Productivos”, la reducción de tiempo promedio de atención “TPA” y el aumento del “NPS” en al menos un 5% y el dashboard operativo, con actualizaciones automáticas y alertas tempranas.

Finalmente, en esta última fase se desarrollaron las estrategias a implementar de acuerdo con los resultados de predicción del modelo, que va a permitir la detección temprana de riesgos y la medición continua de resultados.

Conclusiones

El modelo de aprendizaje automático (Machine Learning) supera el umbral de rendimiento sólido para todas las clases, lo que indica que tiene buena capacidad para predecir casos reales con una macro precisión de 82% y una macro recall del 79%, con estos resultados tenemos un árbol de decisión capaz de predecir la productividad del canal de ventas presencial de una empresa de telecomunicaciones.

La clase “Aceptable” muestra el menor F1-Score al 70%, lo que puede ser un punto de mejora, esto indica un buen equilibrio entre precisión y recall reflejando que el modelo es capaz de identificar adecuadamente los diferentes perfiles sin generar una cantidad excesiva de errores.

La presencia de “NPS” como una variable importante indica que la satisfacción del cliente podría estar correlacionada con la productividad del asesor, lo que refuerza la idea de que un buen servicio se traduce en mejores resultados comerciales.

Las variables más significativas e influyentes del modelo son “Pospago” y “Hogar” lo que impacta directamente al rendimiento del asesor, el bajo rendimiento en estas variables puede deberse a diferencias en la complejidad del servicio y también a la estructura de metas comerciales.

Finalmente, el modelo de productividad se encuentra en proceso de implementación dentro del canal presencial de ventas, en conjunto con otras áreas como Gestión Humana, Comercial y Soporte. Este trabajo permite estructurar estrategias conjuntas que potencien el impacto del modelo, facilitando la toma de decisiones basadas en datos. Asimismo, se ha definido un ciclo de retroalimentación continua que contempla el reentrenamiento mensual del modelo, garantizando su sostenibilidad y adaptabilidad frente a los cambios en el comportamiento de los asesores.

Recomendaciones

Implementar las propuestas de estrategia basadas en los resultados del modelo, inicialmente se debe disponibilizar un dashboard para seguimiento en tiempo real de los indicadores y variables clave que afectan la productividad de los asesores, con el fin de implementar planes de acción que mejoren el desempeño de los asesores que no llevan buenos resultados.

El esquema de incentivos diferenciados a cada clase, estos tienen un impacto tanto individual como colectivo y tiene muy buenos resultados, incentiva el desarrollo y crecimiento de los asesores y el resultado se ve reflejado durante el tiempo que este en vigencia el incentivo.

A futuro el modelo se puede escalar e integrar a sistemas CRM para generación de alertas en tiempo real, asegurando el reentrenamiento del modelo y refinando variables para obtener resultados más precisos.

Focalizar acciones de mejora donde las predicciones del árbol clasifican “Aceptable” y “Retroalimentación”, clases que tienen un gran potencial de mejora en sus indicadores, especialmente en el “NPS” y “Turnos”.

Implementar un piloto en un porcentaje de asesores usando el dashboard para medir la variación de la productividad de un mes a otro, validar métricas del cambio y depende de los resultados obtenidos hacer la implementación a toda la operación comercial.

Bibliografía

- Asif, D., Arif, M. S., & Mukheimer, A. (2025). A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry. *Results in Engineering*, 26, 104629. <https://doi.org/10.1016/j.rineng.2025.104629>
- Bases de Datos*. (s. f.). Recuperado 8 de mayo de 2025, de <https://libros.metabiblioteca.org/items/33223a6a-b4fc-4d78-8f32-6debec4e60cd>
- Bobadilla, J. (2021). *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. Ediciones de la U.
- Cai, A., Liu, M., & Liu, H. (2022). A methodology for evaluating salespeople performance considering efficiency and effect: A case study of a liquor company in China. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.923198>
- Congreso de Colombia, Ley 1581, No. Por la cual se dictan disposiciones generales para la protección de datos personales, Congreso de Colombia (2012). <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- Crea un Arbol de Decisión en Python | Aprende Machine Learning*. (s. f.). Recuperado 9 de mayo de 2025, de <https://www.aprendemachinelearning.com/arbol-de-decision-en-python-clasificacion-y-prediccion/>
- Hussien, A., Maksoud, A., Al-Dahhan, A., Abdeen, A., & Baker, T. (2025). Machine learning model for predicting long-term energy consumption in buildings. *Discover Internet of Things*, 5(1), 18. <https://doi.org/10.1007/s43926-025-00115-7>
- Jaber, J., Kasasbeh, B., AlSawareah, W., Qaddoura, R., & Kamal, S. (2023). Customer Complaint Characteristics and Predictive Modeling at Orange Telecom: A Machine Learning Analysis. *2023 2nd International Engineering Conference on Electrical,*

Energy, and Artificial Intelligence (EICEEAI), 1-6.

<https://doi.org/10.1109/EICEEAI60672.2023.10590245>

Jiménez, E. C. (2021). *Introducción al Machine Learning con MATLAB*. Marcombo.

Laguna, C. (s. f.). *4. Correlación y Regresión lineal*.

Lin, X., Lu, T.-J., & Chen, X. (2019). Total factor productivity growth and technological change in the telecommunications industry. *Discrete and Continuous Dynamical Systems - Series S*, 12(4-5), 795-809. Scopus. <https://doi.org/10.3934/dcdss.2019053>

Liu, Y., Li, K., Zhang, Y., Cai, Y., Liu, X., Jia, Y., Yao, P., Wei, X., Wu, H., Liu, X., Feng, C., Li, C., Wang, W., Zhang, S., & Du, C. (2025). Impact of phthalate exposure and blood lipids on breast cancer risk: Machine learning prediction. *Environmental Sciences Europe*, 37(1), 39. <https://doi.org/10.1186/s12302-025-01071-3>

Martínez, R. E. B., Ramírez, N. C., Mesa, H. G. A., Suárez, I. R., León, P. P., & Morales, S. L. B. (2009). *Árboles de decisión como herramienta en el diagnóstico médico*.

Mekdad, D. K., Jayasuria, J. G., & Hossin, A. A. (2024). Impact of training and development on employee performance: Insights for sustainable capacity building of employees in telecommunication organization in Qatar. En *Innovation and Technological Advances for Sustainability*. CRC Press.

Productividad y competitividad. (s. f.).

¿Qué es la minería de datos? La minería de datos, explicada - AWS. (s. f.). Amazon Web Services, Inc. Recuperado 8 de mayo de 2025, de <https://aws.amazon.com/es/what-is/data-mining/>

Ramkumar, M., Malathi, K., & Pavithra, K. (2023). Optimizing Machine Learning Model Accuracy via OBNT Algorithm: Advanced Data Preprocessing Technique. *2023*

International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 1-6.

<https://doi.org/10.1109/ICSES60034.2023.10465344>

Soto, C., & Jiménez, C. (2011). APRENDIZAJE SUPERVISADO PARA LA DISCRIMINACIÓN Y CLASIFICACIÓN DIFUSA. *DYNA*, 78(169), 26-33.