Modelo de clasificación predictiva de renovación: Caso de estudio en una empresa funeraria

Gloria Patricia Cadavid Zuluaga
Laura Hernández Rebolledo

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor
Ronald Akerman Ortiz García, Magíster en Ingeniería

Universidad de Antioquia
Facultad de Ingeniería
Especialización en Analítica y Ciencia de Datos
Medellín, Antioquia, Colombia
2025

Cita	(Cadavid Zuluaga & Hernández Rebolledo, 2025)
Referencia **Estilo APA (2020)**	Cadavid Zuluaga, G., & Hernández Rebolledo, L. A. (2025). *Modelo de clasificación de clientes para la unidad de negocio de previsión en una empresa funeraria*. Universidad de Antioquia, Medellín, Colombia.

Especialización en Analítica y Ciencia de Datos, Cohorte VIII.

Centro de Investigación Ambientales y de Ingeniería (CIA).

UNIVERSIDAD DE ANTIOQUIA
Vicerrectoría de Docencia

Sistema de Bibliotecas

Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co
Rector: John Jairo Arboleda Céspedes.
Decano: Julio Cesar Saldarriaga Molina
Jefe departamento: Danny Alexandro Múnera Ramírez
Coordinadora del Programa: Maria Bernarda Salazar Sánchez

Tabla de contenido

Lista de Figuras

Lista de Tablas

Siglas, acrónimos y abreviaturas

IEEE	Instituto de Ingenieros Eléctricos y Electrónicos
ML	Machine Learning
UdeA	Universidad de Antioquia
MLP	Multi Layer Perceptron
CSV	Coma separeted values
SQL	Structure Query Language
AUC	Area under curve
DIAN	Dirección de impuestos y aduanas Nacional
EDA	Exploratory data análisis
CART	Classiffication and Regression Trees
ANN	Artifitial neuronal networks
SGD	Stochastic Gradient Descent
ROC	Receiver Operating Characteristic

Modelo de clasificación predictiva de renovación: Caso de estudio en una empresa funeraria

***Resumen*— En el presente trabajo se implementan metodologías de clasificación supervisada para predecir si un cliente renueva o no su cobertura al servicio de asistencias de la línea de negocio de Previsión en la Previsora Social Cooperativa Vivir, ya que impacta directamente en la sostenibilidad del negocio a largo plazo. De ahí que, el objetivo de este trabajo es implementar métodos de machine learning para identificar patrones y características relevantes de los clientes, tal que se pueda segmentar y priorizar los esfuerzos de retención de manera más eficiente. De esta forma, se optimizarán los recursos en las campañas de fidelización y se maximizará el impacto de las acciones comerciales. El modelo también brindará insights valiosos sobre los factores que influyen en la renovación de contratos, facilitando decisiones más informadas y estrategias más efectivas para mejorar la tasa de renovación y garantizar la sostenibilidad del negocio a largo plazo. Se utilizan metodologías cómo árboles de decisión, Ramdom Forest y Redes Neuronales con la optimización de hiperparámetros. Se obtuvieron resultados que permiten realizar la correspondiente clasificación con un alto índice precisión.**

***Palabras clave:* Renovación, Clasificación, Asignación de recursos, Machine learning.**

1. Introducción

En los últimos años, el sector asegurador en América Latina ha mostrado una evolución notable impulsada por la creciente demanda de servicios de protección y bienestar integral para las familias [1]. Dentro de este panorama, las coberturas de previsión exequial han ganado protagonismo como una alternativa accesible que permite a las personas planear con anticipación la atención funeraria y otros servicios de asistencia. Esta transformación del mercado también ha alcanzado al sector funerario, donde organizaciones como Previsora Social Cooperativa Vivir han ampliado su oferta de servicios hacia modelos más integrales, enfocados en la atención continua y la fidelización de sus afiliados.

En este contexto, la ciencia de datos emerge como una herramienta estratégica para afrontar los retos del sector, especialmente frente a problemáticas como la cancelación de contratos de previsión. [2] A pesar de ofrecer precios asequibles y servicios de alto impacto emocional y económico, muchas empresas funerarias enfrentan tasas elevadas de cancelaciones, lo que afecta directamente la sostenibilidad del negocio. Uno de los principales desafíos actuales es anticipar el comportamiento del cliente, en particular, identificar a tiempo aquellos perfiles con mayor o menor posibilidad de renovar su contrato, para así enfocar mejor los esfuerzos de retención y optimizar el uso del presupuesto de mercadeo [3]. Todo lo anterior, dado que actualmente la tasa de deserción es alta y tiene implicaciones en el negocio.

Es así como desde la analítica de datos, se ha evidenciado que la implementación de metodologías de clasificación de machine learning, ofrecen una solución para este tipo de contextos [4] [5]. Dichos modelos han sido ampliamente utilizados en sectores como financiero, de seguros [6] para mejorar y optimizar la toma de decisiones, específicamente las aplicaciones en las empresas de seguros. Sin embargo, el alcance de dichos modelos es nuevo para este negocio, dado que si bien es semejante no tiene el mismo comportamiento, ni tipo de clientes.

Este trabajo busca implementar y evaluar modelos de clasificación supervisados para optimizar la clasificación de renovación, contribuyendo así a mejorar la eficiencia operativa y toma de decisiones del área de marketing en la definición de sus estrategias. Esto no solo aumentaría la tasa de conversión de dichas campañas, sino que también permitiría focalizar esfuerzos en perfiles de clientes con alta probabilidad de renovar sus coberturas [7], lo cual contribuye directamente al crecimiento sostenido y la estabilidad del negocio a largo plazo. En el desarrollo de este proyecto se utilizaron registros históricos anonimizados, provenientes de las bases de datos internas de la organización, que incluyen variables como edad, ubicación, tipo de cobertura contratada, frecuencia de pagos, y registros de contacto. Sobre esta

base, se implementaron técnicas de aprendizaje automático supervisado como árboles de decisión, Random Forest, y Redes neuronales, evaluando su desempeño mediante métricas estándar como Matriz de confusión, exactitud, precisión, recall y AUC [8] [9].

A continuación, se describen a detalle las diferentes metodologías utilizadas, con su correspondiente soporte académico, evidenciando el comportamiento de cada metodología implementada con sus correspondientes resultados.

2. Materiales y Métodos

2.1. Descripción del conjunto de datos

Los datos utilizados en este estudio provienen de la base de datos transaccional interna de Previsora Social Cooperativa Vivir, y corresponden a los clientes activos del Territorio de Previsión durante el año 2023. Fueron extraídos directamente mediante consultas SQL y exportados en formato CSV, con un tamaño total de 18,6 MB. El conjunto de datos se construyó a partir de la unión de varias tablas relacionales, e incluye variables como edad, número de pagos, mecanismo de pago, composición del producto actual y zona geográfica (ver Figura 1). En total, se recopilaron 184,014 registros correspondientes a titulares de contrato y 21 variables, con datos tanto cuantitativos como cualitativos. Estos datos en su mayoría son datos discretos y categóricos. Por la naturaleza del negocio.

#	Variable	Dtype
0	Canal	object
1	Estado	object
2	Edad_titular	float64
3	Antiguedad_titular	int64
4	Forma_pago	int64
5	Numero_de_pagos	int64
6	Valor_Pagado	Int64
7	Contratos_activos_anteriores	int64
8	Valor_Contratos_activos_anteriores	int64
9	Casado	int64
10	Beneficiario	int64
11	Mascotas	int64
12	Exequial_Personas	int64
13	Auxilio	int64
14	Asistencia_Persona	int64
15	Total_prima_mes	Int64
16	tipo_geografia	object
17	Tipo_pago	object
18	Mes_renovacion	int32

Figura 1. Descripción, tipo de variables

Así mismo, para garantizar la protección de la información, solo se incluyen dos variables sensibles (edad y género), y se anonimiza mediante un Id Titular generado aleatoriamente. El acceso al archivo se encuentra restringido mediante token y se gestiona desde un repositorio privado, asegurando su integridad y confidencialidad. Cabe mencionar que, durante el tratamiento de los datos, se evidenciaron muchos datos faltantes para en la variable género. Inicialmente, se implementó un modelo predictivo para la imputación de la variable, basada en los nombres, sin embargo, al realizar el análisis exploratorio de los datos, se identificó que dicha variable no tiene un aporte significativo para la variable de salida, de forma que se toma la decisión de eliminarla de los datos.

A continuación, se hará una descripción del preprocesamiento y limpieza de los datos, análisis exploratorio y sus hallazgos. Se hará uso de las técnicas, arboles de decisión, random forest y redes neuronales combinando una capa convolucional con una capa densa convencional.

Teniendo en cuenta la exploración inicial de los datos, se evidencia que existe un desbalanceo de los datos (ver Figura 2). De ahí que se deberán implementar medidas para el balanceo de las clases de forma que no se minimice el riesgo de un sobreajuste al momento de implementar las técnicas de modelado.

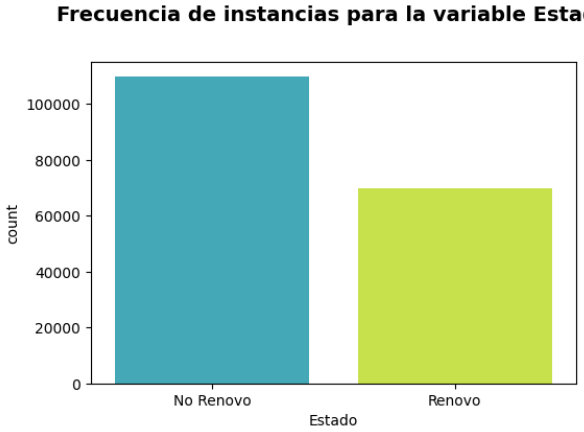

Figura 2. Distribución de la variable de estado

2.2. Procesamiento y limpieza de los datos

2.2.1. Escalado y normalización

Para homogenizar las variables numéricas y facilitar su comparación, se aplicó la técnica de estandarización utilizando la función StandardScaler de la biblioteca Scikit-learn. Este método transforma los datos para que tengan una media de cero y una desviación estándar de uno, lo cual es especialmente útil cuando las variables presentan diferentes escalas y unidades.

Además, se incorporó una base de datos adicional proporcionada por la Dirección de Impuestos y Aduanas Nacionales (DIAN) en formato CSV, que contiene 727 registros clasificados por tipo de zona (rural o urbana). Con el objetivo de reducir la dimensionalidad y preparar los datos para el modelado, se aplicaron técnicas de codificación:

- **Codificación One-Hot:** Utilizada para variables categóricas nominales, como el tipo de geografía, donde no existe un orden inherente entre las categorías.
- **Codificación de Etiquetas (Label Encoding):** Aplicada a variables categóricas ordinales, como el mecanismo de pago, donde las categorías tienen un orden lógico.
- **Codificación Ordinal:** Implementada en el segmento de venta, jerarquizando las categorías según la importancia o preferencia empresarial.

2.2.2. Cálculo y eliminación de valores atípicos

Para identificar y eliminar valores atípicos en las variables numéricas, se empleó el algoritmo de los k vecinos más cercanos (k-NN). El valor óptimo de k se determinó mediante una búsqueda de hiperparámetros utilizando la técnica de Grid Search, resultando en k=5. Este enfoque permite detectar observaciones que difieren significativamente de sus vecinos más cercanos, indicando posibles anomalías.

2.3. Análisis exploratorio

Se realizó un análisis exploratorio de datos (EDA) para comprender la distribución y balanceo de las variables tanto categóricas como numéricas. Se observaron desequilibrios en variables como el segmento de venta y el tipo de geografía, con una predominancia de datos en zonas urbanas. Y en ese mismo sentido la cantidad de muestras que tienen forma de pago entre 30 y 360 días. Se observa que la variable tener mascota tiene una incidencia directa en el valor de la prima (ver Figura 3). Además, se identifica que el número de renovaciones es inversamente proporcional al modo de pago, considerando este factor como una característica del contrato. Solo el producto de previsión y la prima tienen una relación directa, mientras que las coberturas de auxilio y asistencia se toman principalmente como adicionales.

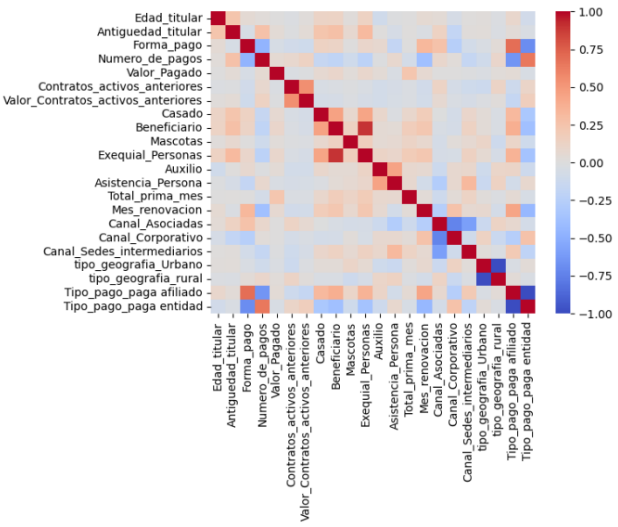

Figura 3. Análisis de correlación entre las variables de la base de datos

También se nota que el número de pagos es inversamente proporcional a la asistencia. Este hallazgo sugiere un aspecto negativo, ya que la inclusión de asistencia puede influir en la decisión de si se realiza o no el pago. Finalmente, este análisis permite descartar variables adicionales que no aportan valor para la capacidad predictiva del modelo. En la Figura 4, se pueden evidenciar el comportamiento de las variables de acuerdo a al estado o no de renovación.

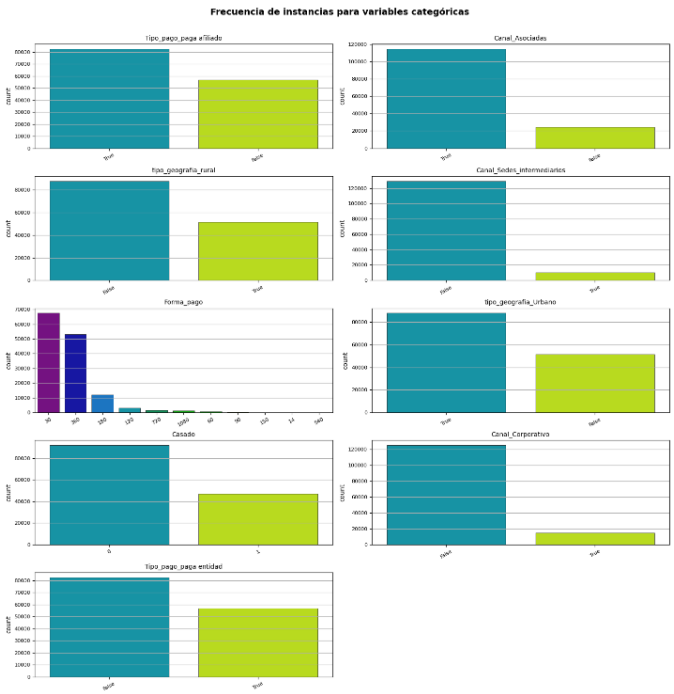

Figura 4. Frecuencias de instancias para variables categóricas

Es evidente el desbalance presente en los datos, dada la naturaleza del negocio, se utiliza el método de resample para ajustar las clases equidistribuidas para la

aplicación de las diferentes metodologías, garantizando que, desde los datos, al momento de aprender no existan riegos de sobreajuste.

Variable	Correlación
Forma_pago	0.40
Exequial_Personas	0.23
Casado	0.19
Auxilio	0.11
Mascotas	0.08
Valor_Pagado	0.06
Edad_titular	0.05
Total_prima_mes	0.03
Asistencia_Persona	-0.03
Antiguedad_titular	-0.05
Contratos_activos_anteriores	-0.13
Valor_Contratos_activos_anteriores	-0.13
Numero_de_pagos	-0.46

Figura 5. Correlación de variables numéricas

2.4. Entrenamiento de modelos

2.4.1. Árboles de decisión

Los clasificadores basados en árboles de decisión son algoritmos de aprendizaje supervisado no paramétrico que tienen como propósito asignar una clase a una instancia, utilizando como base uno o varios atributos conocidos del conjunto de datos. Esta técnica destaca por su capacidad interpretativa y su facilidad de implementación, lo que la ha posicionado como una de las más utilizadas en problemas de clasificación y regresión.

La metodología se fundamenta en un proceso de partición recursiva del espacio de atributos, construyendo reglas de decisión en forma de estructura arbórea. Una implementación ampliamente reconocida de esta técnica es el algoritmo de Clasificación y Regresión (CART) en la Figura 6 se evidencia la estructura de formación de un árbol de decisión, el cual construye árboles binarios seleccionando atributos según el índice de Gini como criterio de impureza. Este enfoque permite dividir los datos de forma eficiente, maximizando la pureza de los subconjuntos resultantes. A diferencia de otros métodos, CART no impone reglas de parada durante el crecimiento del árbol; en cambio, permite que el árbol se expanda completamente y luego aplica una técnica de poda para evitar el sobreajuste, preservando así patrones significativos que podrían perderse si el proceso de partición se detuviera prematuramente.

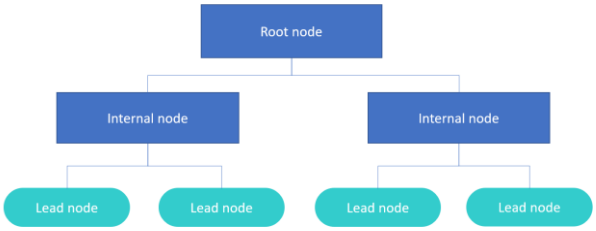

Figura 6. Metodología de formación del algoritmo árbol de decisión [10]

2.4.2. Random Forest

El algoritmo Random Forest es una técnica de aprendizaje supervisado basada en el enfoque de ensamblado (ensemble learning), cuyo objetivo principal es construir un modelo predictivo robusto a partir de la combinación de múltiples clasificadores individuales, en este caso, árboles de decisión. Este enfoque busca reducir tanto el sesgo como la varianza del modelo resultante, aprovechando la diversidad generada por múltiples árboles entrenados sobre subconjuntos aleatorios del conjunto de datos y características.

Random Forest, propuesto por Breiman, funciona mediante la generación de un número elevado de árboles de decisión, cada uno construido sobre una muestra aleatoria del conjunto de entrenamiento mediante el método de bootstrap (muestreo con reemplazo). [9]Además, en cada división de un nodo, se selecciona aleatoriamente un subconjunto de predictores, lo que introduce mayor diversidad entre los árboles. El resultado final del modelo se obtiene mediante votación mayoritaria en clasificación o

promedio en regresión, lo que incrementa la estabilidad y precisión del modelo.

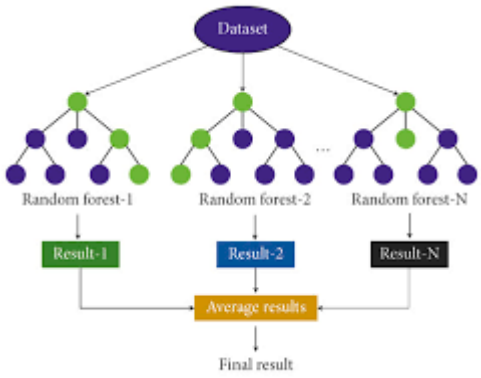

Figura 7. Metodología de formación del algoritmo Random Forest [11]

La construcción de los árboles se realiza en dos etapas principales: primero, se selecciona aleatoriamente un subconjunto de observaciones para entrenar cada árbol; segundo, en cada nodo del árbol se evalúa un subconjunto aleatorio de características para definir la mejor partición. Los árboles crecen sin poda, lo que permite capturar patrones complejos en los datos, mitigando el sobreajuste mediante la agregación de múltiples predicciones.

2.4.3. Redes neuronales

Las redes neuronales artificiales (ANN, por sus siglas en inglés) son modelos computacionales inspirados en el funcionamiento del cerebro humano, diseñados para identificar patrones complejos mediante aprendizaje supervisado. Estas redes aprenden a partir de ejemplos, ajustando sus pesos internos a medida que se les expone a más datos, lo cual incrementa su capacidad predictiva con el tiempo y permite mejorar su precisión mediante múltiples iteraciones de entrenamiento.

En este proyecto se implementa una variante ampliamente utilizada conocida como Perceptrón Multicapa (MLP, por sus siglas en inglés), que se caracteriza por tener una arquitectura compuesta por una capa de entrada, una o más capas ocultas y una capa de salida. Cada capa está conformada por nodos (neuronas artificiales) que están completamente conectados entre sí. Estas capas ocultas son responsables de realizar las transformaciones no lineales necesarias para que la red aprenda funciones complejas.

Aunque las redes neuronales pueden aplicarse tanto a tareas supervisadas como no supervisadas, en este caso el enfoque es completamente supervisado, es decir, la red aprende a partir de ejemplos etiquetados en los que se conoce la clase o resultado esperado.

Adicionalmente, considerando el tipo de datos, y del comportamiento de la red artificial, el cual no fue el esperado. Se ejecutó una arquitectura anidada, combinando una capa convolucional, con el objeto de identificar patrones en los datos, y con base en esos patrones aprendidos entregarlos a una capa densa, entregando a la red, información organizada para facilitar el aprendizaje de patrones y generalizaciones.

La Figura 8 representa una simulación extendida a la arquitectura implementada, dado que inicialmente se ejecuta una capa convolucional y posterior una capa densa para el proceso de clasificación.

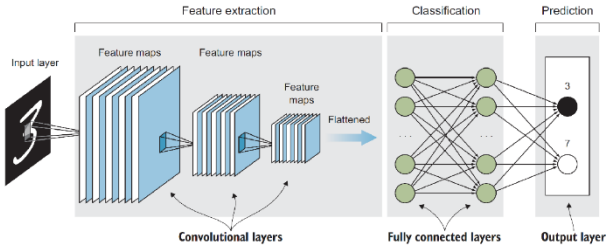

Figura 8. Metodología de formación del algoritmo Redes neuronales convencionales [12]

Durante el entrenamiento, se implementan técnicas de optimización como la propagación hacia atrás (backpropagation) y el uso de algoritmos como Adam o SGD (Stochastic Gradient Descent) para minimizar la

función de pérdida. Además, se puede emplear validación cruzada y ajuste de hiperparámetros mediante Grid Search o Random Search para mejorar la arquitectura y el rendimiento del modelo.

2.5. Métricas de evaluación

En general para evaluación del rendimiento predictivo de los modelos, se emplean métricas estándar como la precisión (precision), la sensibilidad o exhaustividad (recall) y la medida F1 (F1 Score), que proporcionan una visión integral del equilibrio entre falsos positivos y falsos negativos en contextos de clasificación desbalanceada. Además, se hace uso de la matriz de confusión para analizar el poder de predicción según las dos clases analizadas.

En cuanto a la optimización de los modelos, se aplicó una búsqueda sistemática de hiperparámetros mediante la técnica de Grid Search, que permite explorar combinaciones de parámetros relevantes (como la profundidad de los árboles, el número mínimo de muestras por hoja, etc.), con el objetivo de maximizar el rendimiento del modelo en conjuntos de validación cruzada.

Para la evaluación del rendimiento del modelo, se emplean métricas comúnmente utilizadas en problemas de clasificación, como precisión (precision), recall (exhaustividad) y la medida F1 (F1 score), las cuales permiten evaluar la capacidad del modelo para identificar correctamente las clases en presencia de desequilibrios en los datos, proporcionando un balance entre exactitud y exhaustividad [5]. Asimismo, se incluye la curva ROC (Receiver Operating Characteristic), que permite evaluar visualmente la capacidad de discriminación del modelo entre las clases, mediante el análisis de la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos.

3. Resultados y Discusión

Inicialmente, se implementaron cuatro modelos base utilizando dos enfoques de aprendizaje automático: árboles de decisión y redes neuronales. El objetivo de esta etapa fue evaluar el desempeño de cada modelo en la clasificación de la renovación de contactos, considerando métricas clave como Accuracy, F1-score y Recall.

En la Tabla 1, se presentan los resultados obtenidos para cada modelo, lo que permitirá analizar su precisión, capacidad de generalización y eficacia en la detección de patrones. Teniendo en cuenta los resultados obtenidos, se puede evidenciar especialmente, tanto para los árboles de decisión, cómo para los random forest, que tienen una buena capacidad predictiva, con resultados de precisión y recall entre el 97% y el 98% de forma correspondiente, los cuales tienen mucho mejor comportamiento comparados con los resultados de las métricas del MLP, con un ROC-AUC del 51%.

Tabla 1. Métricas obtenidas para los diferentes modelos analizados

Modelo	Accuracy	F1	Recall
Random Forest	0,98	0,98	0,97
Arboles de decisión	0,97	0,97	0,97
MLP	0,51	0,67	0,75
Convolucionadas - MLP	0,98	0,98	0,98

Además del análisis de indicadores de desempeño, se evaluaron las matrices de confusión de cada modelo para obtener una visión más detallada sobre sus errores y aciertos en la clasificación. Los modelos Random Forest (Figura 10), Árbol de Decisión (Figura 11) y Convolucional-MLP (Figura 12) presentan una

distribución equilibrada de aciertos en verdaderos positivos y verdaderos negativos, minimizando errores en la clasificación. En contraste, el modelo MLP evidencia una alta tasa de falsos positivos, lo que indica dificultades para discriminar correctamente los casos negativos.

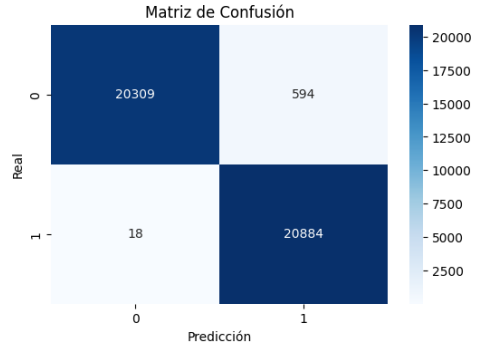

Figura 9. Matriz de confusión Random forest

En contraste, el uso de técnicas de optimización de hiperparámetros cómo el Gridsearch, si bien evidencia cierta mejora en las métricas, no necesariamente cómo mecanismo optimizador, tiene un impacto significativo sobre la capacidad predictiva de dichos modelos.

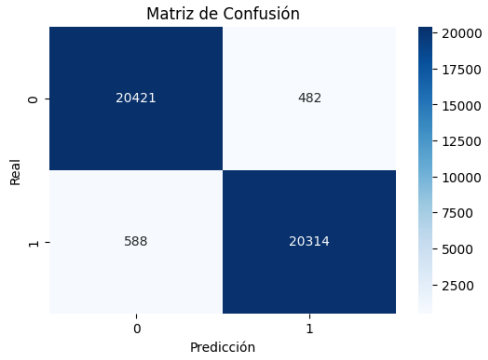

Figura 10. Matriz de confusión árbol de decisión

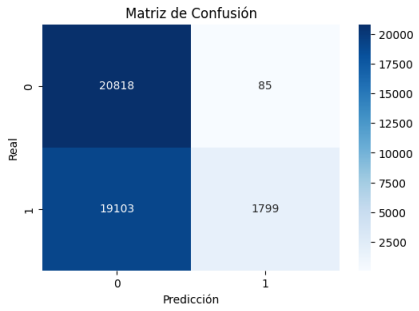

Figura 11. Matriz de confusión redes neuronales convencionales

Por otro lado, si bien se hicieron esfuerzos por optimizar las métricas y el proceso de entrenamiento del modelo MLP, con una arquitectura ajustada (Figura 13), no hay mejora con el cambio de los hiperparámetros con respecto a los resultados de los demás modelos.

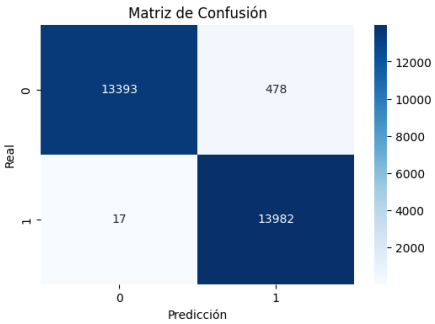

Figura 12. Matriz de confusión Redes neuronales convolucionadas 1 dimensión unida a una red convencional

Así mismo, considerando la bibliografía, si bien los modelos de redes neuronales clásicos MLP, pueden funcionar para la clasificación de clases binarias, no necesariamente aplicar para el problema de negocio abordado, esto puesto que puede deberse a la necesidad de grandes volúmenes de datos para el proceso de entrenamiento y aprendizaje de patrones para la red. Otra posibilidad de limitaciones con respecto a la implementación de la red es la profundad de la misma, o incluso la arquitectura, puesto que no tiene una gran profundidad por lo que, de acuerdo con los resultados para la quinta época, ya habría alcanzado su capacidad máxima de aprendizaje.

Considerando los resultados obtenidos, se entiende entonces que ambos modelos, árboles de decisión y random forest, se entrenaron con suficientes datos y aprendieron de forma que generalizan con una precisión de un 98%. Y esto implica que los modelos con capaces de predecir tanto los clientes que van a renovar, cómo los que no van a renovar, minimizando el riesgo de sobreajuste.

Dado que la implementación de este tipo de técnicas no es frecuente para entornos de empresas de Previsión funeraria. La comparación de los resultados con respecto a otro tipo de técnicas aplicadas no es posible, sin embargo, si es evidencia tanto desde la bibliografía, cómo desde la práctica, que, considerando las características particulares de los datos, los modelos más apropiados para el abordaje de la problemática son los que mejores resultados arrojaron.

Ahora, considerando los resultados obtenidos, una de las métricas más utilizadas para la decisión de cuál de los modelos es el apropiado, es la precisión, sin embargo, se define cómo métrica decisiva el F1 score, dado que los resultados de la precisión son variables para las diferentes métricas y pueden representar sesgo en los resultados del modelo, dado que toma como base la capacidad predictiva con respecto a una de las 2 clases, de ahí que con el F1 score permite tomar decisiones más balanceadas con respecto a la generalización en la clasificación de los datos.

Así, para el caso en particular, el F1 score en todos los modelos es del 98%, al presentarse esta situación y teniendo en cuenta que el objetivo final es entregar una herramienta escalable al negocio. Una de las métricas del negocio para decidir de cuál de los modelos elegir, es la practicidad en la implementación del modelo elegido. De ahí que, entre los 3 modelos con mejor comportamiento y facilidad en la implementación, se concluye que el Random Forest, representa una buena aproximación desde la analítica de datos, para dar solución a la problemática planteada.

4. Conclusiones

- La experimentación con metodologías como árboles de decisión y Random Forest, aportan escalabilidad y adaptación en el aprendizaje de patrones. dadas las características de los datos del problema empresarial.
- Por el origen y tipología de los datos, se evidenció que son de alta transaccionalidad, siendo más eficiente la identificación de patrones a través de una capa convolucional y posterior el entrenamiento de una capa densa para lograr una generalización más acertada con modelos de redes neuronales.
- Se comprueba que la implementación de aprendizaje automático para la solución de problemas empresariales no necesariamente exige grandes inversiones en la infraestructura tecnológica, dado a que con técnicas sencillas obtenemos resultados que aportan a la solución
- Se logran entrenar modelos que generalizan el comportamiento de los clientes, permitiendo la identificación de patrones, aportando perspectivas diferentes para la toma de decisiones en el negocio, específicamente para la segmentación de los clientes y las diferentes estrategias de marketing.

Referencias

[1] Fundación MAPFRE, «Tendencias del mercado de seguros en América Latina para 2024,» Fundación MAPFRE, Madrid, 2024.

[2] M. S. Sardaña, «Modelo predictivo de venta cruzada en productos de Vida y Salud: Random Forest vs XGBoost,» Universidad Carlos III de Madrid, Madrid, 2022.

[3] S. D. A. B. A. D. a. R. R. Prof. Nikhil Patankar, «ReserchGate,» 12 2021. [En línea]. Available: https://www.researchgate.net/publication/356756320_Customer_Segmentation_Using_Machine_Learning. [Último acceso: 01 5 2025].

[4] A. K. N. Y. R. A. Shashi Pal Singh, «ReserchGate,» 18 5 2018. [En línea]. Available: https://www.researchgate.net/publication/339562150_Data_Mining_Consumer_Behavior_Analysis. [Último acceso: 1 5 2025].

[5] L. T. M. R. Carolina Alvarez Florez, «Modelos de aprendizaje automático para la predicción de la preferencia en el uso de canales de atención para un fondo de pensiones y cesantías,» Universidad de Antioquia, Medellín, 2021.

[6] S. S. H. G. C. M. Payam Boozary, «Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction,» International Journal of Information Management Data Insights, vol. 5, nº 1, p. 15, 2025.

[7] E. C. C.-E. D. B. Bolaji Iyanu Adekunle, «Improving Customer Retention Through Machine Learning : A Predictive,» International Journal of Scientific Research in Computer Science, Engineering and, vol. 9, nº 4, pp. 507-523, 2023.

[8] T. I. a. S. Adewale, «Implementation of Data Mining on a Secure Cloud Computing over a Web API using Supervised Machine Learning Algorithm,» International Journal of Advanced Computer Science and Applications, vol. 13, nº 5, pp. 112-119, 2022.

[9] J. H. Friedman, «Greedy Function Approximation: A Gradient Boosting Machine,» The Annals of Statistics, vol. 29, nº 5, pp. 1189-1232, 2001.

[10] IBM, «IBM,» IBM, 2025. [En línea]. Available: https://www.ibm.com/es-es/think/topics/decision-trees. [Último acceso: 11 Abril 2025].

[11] ReserchGate, «ReserchGate,» ReserchGate, 2024. [En línea]. Available: https://www.researchgate.net/figure/Flowchart-of-random-forest-algorithm_fig1_358926904. [Último acceso: 11 Abril 2025].

[12] J. A. J. Pacheco, «Medium,» Medium, 30 Noviembre 2023. [En línea]. Available: https://medium.com/@jajp2203/clasificaci%C3%B3n-de-im%C3%A1genes-con-redes-profundas-a229774e53f4. [Último acceso: 11 Abril 2025].

[13] S. F. Sabbeh, «Machine-Learning Techniques for Customer Retention: A Comparative Study,» International Journal of Advanced Computer Science and Applications, vol. 9, nº 2, p. 10, 2018.

[14] F. L. Aurora, «Funerales La Aurora,» 04 Enero 2022. [En línea]. Available: https://funeraleslaaurora.com/2022/01/04/efectos-de-la-pandemia-la-mirada-desde-una-funeraria/. [Último acceso: 6 Noviembre 2024].