

# Implementación de modelos de Inteligencia Artificial para la retención de clientes en la industria de telecomunicaciones

Aldo Daniel Ojeda Rodríguez, *Universidad Autónoma de Nuevo León, Facultad de Ciencias Físico Matemáticas, Maestría en Ciencia de Datos.*

**Resumen**—La retención de clientes es un factor sumamente importante siendo un problema que más afectan a la industria de telecomunicaciones, por lo cual, aplicaremos algoritmos de inteligencia artificial con el objetivo de poder clasificar la retención de clientes entre aquellos que son retenidos y no retenidos, además de seleccionar las variables más importantes para un modelo de inteligencia artificial supervisado.

## I. INTRODUCCIÓN

Dentro de la industria de telecomunicaciones uno de los problemas que más afectan a la industria es el de la retención de clientes, es más costoso conquistar clientes nuevos que mantener a aquellos que ya son consumidores de nuestro servicio, por lo cual las empresas están buscando generar algoritmos capaces de predecir cuando un cliente es propenso al abandono [1].

En el presente trabajo compararemos diversas técnicas de clasificación y evaluaremos aquella que sea la mejor para la clasificación de retención de clientes (Churn).

## II. MARCO TEÓRICO

### II-A. Modelos No Supervisados

El algoritmo de agrupación  $K$ -medias separa el conjunto de datos de entrada de  $N$  filas en un número menor de subconjuntos  $K$ , siendo  $K < N$ . El parámetro  $K$ , se elige aleatoriamente, enseguida se calcula la media de cada subconjunto de datos y cuantifica la distancia entre ellos repitiendo estos pasos hasta generar el centroide óptimo [2], de acuerdo a la ecuación 1

$$J = \sum_{i=1}^m \sum_{k=1}^K w_{ik} |x_i - \mu_k|^2, \quad (1)$$

donde  $w_{ik} = 1$  para el punto de datos  $x_i$  si pertenece al subconjunto  $k$ ; en caso contrario,  $w_{ik} = 0$ . Además,  $\mu_k$  es el centroide de  $x_i$ .

En los métodos  $K$ -medoides, un conglomerado está representado por uno de sus puntos. Cuando se seleccionan los medoides, los conglomerados se definen como subconjuntos de puntos cercanos a los medoides respectivos, y la función objetivo se define como la distancia promediada u otra medida de disimilitud entre un punto y su medoide [3].

siendo la función objetivo presentada en la ecuación 2

$$J = \sum_{n_i} \sum_{m_i \in n_i} |N_i - M_i|. \quad (2)$$

Por último, el Agrupamiento jerárquico construye una jerarquía de conglomerados también conocido como dendrograma. Los métodos de agrupación jerárquica se clasifican en métodos de aglomeración (ascendentes) y métodos divisivos (descendentes). Una agrupación de aglomeración comienza con conglomerados de un punto y fusiona recursivamente dos o más conglomerado más apropiados [4].

### II-B. Modelos Supervisados

La regresión Logística utiliza la función logística nombrada como función Sigmoidal, para mapear predicciones y sus probabilidades, esta función es una función en forma de S, la cual convierte un valor de entrada ( $x$ ) en un rango entre 0 y 1 dado por la siguiente ecuación:

$$f(x) = \frac{1}{1 + \exp(-x)}$$

siendo la regresión logística dada por la ecuación II-B.

$$P(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)},$$

siendo  $\beta_0$  el intercepto del argumento y  $\beta_1$  el coeficiente que acompaña a la variable independiente.

Esta ecuación es similar a una regresión lineal donde las variables de entrada se combinan linealmente para predecir una variable de salida la cual utiliza coeficientes, sin embargo esta regresión permite generar un valor entre 0 y 1 (binaria) [5].

El algoritmo conocido como AdaBoost para una clasificación binaria definimos la función margen de interacción de cada par de entrada-salida () por la ecuación 3,

$$mg(z_i, c_i) = y_i \sum_{t=1}^T c_t h_t(x_i), \quad (3)$$

siendo,  $z_1 = (x_i, y_i), i = 1 \dots N$ ,  $h_t(x_i) = \pm 1$  y  $c_t$  es la variable del peso la cual se busca optimizar maximizando esta función [6].

El algoritmo LGBM es el algoritmo basado en un árbol de decisión el cuál es más rápido que otros algoritmos, ya que progresa verticalmente además de que, al ser un algoritmo

basado en un árbol, tiene una raíz y una hoja que pueden crecer vertical u horizontalmente.

LGBM es beneficioso si estamos calculando resultados en un conjunto de datos grande; de lo contrario, puede ajustarse en exceso a un conjunto de datos pequeño. La principal ventaja de este algoritmo es que es muy ligero y consume muy poca memoria para calcular miles de filas, proporcionando resultados muy precisos [7].

### II-C. transformación de Datos y Métricas de Desempeño

La transformación de Maximo-Mínimo es un tipo de normalización que puede escalar todos los valores de un conjunto de datos para que tengan valores de 0 a 1[8], de acuerdo a la ecuación 4.

$$x_s = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Para la evaluación de los modelos de clasificación hay cuatro términos que deben de ser definidos los cuales son los más utilizados en la literatura [9]:

- TP (Verdaderos positivos): el número de clientes que deberían estar en la categoría de abandono y el algoritmo de predicción ha determinado correctamente su categoría
- TN (Verdaderos negativos): el número de clientes que deberían estar en la categoría de retención y el algoritmo de predicción ha determinado correctamente su categoría
- FP (Falsos positivos): número de clientes que no son retenidos pero que el algoritmo ha clasificado incorrectamente como retenido.
- FN (Falsos negativos): número de clientes que son no retenido pero que el algoritmo ha clasificado incorrectamente como retenido.

exhaustividad es la proporción de clientes que abandonan correctamente, y se calcula mediante la ecuación 5.

$$\text{Exhaustividad} = \frac{TP}{TP + FN} \quad (5)$$

Precisión es la proporción de clientes que abandonan correctamente, y se calcula mediante la ecuación 6.

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (6)$$

Para evaluar el rendimiento de los clasificadores en la predicción de la retención de clientes, utilizamos las medidas de precisión, exhaustividad, exactitud, medida-F y ROC-AUC [10].

La precisión o la exhaustividad por sí solas no pueden describir la eficacia de un clasificador, ya que un buen rendimiento en uno de esos índices no implica necesariamente un buen rendimiento en el otro. Por este motivo, se suele utilizar la medida F, como métrica única para evaluar el rendimiento de los clasificadores, siendo la medida F definida como:

$$F = \frac{2 \times \text{Precisión} \times \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}} \quad (7)$$

La característica operativa del receptor (ROC) y el área bajo la curva (AUC) se utiliza para describir la capacidad de

diagnóstico de un clasificador, que se representa mediante una curva [11]. Si se traza la tasa de verdaderos positivos (TPR) o frente a la tasa de falsos positivos (FPR) con varios ajustes, se obtiene la curva ROC-AUC:

$$FPR = \frac{FP}{(FP + TN)} = 1 - TNR \quad (8)$$

siendo  $FP$  = falsos positivos,  $TN$  = verdaderos negativos y  $FN$  = falsos negativos.

El índice de Silueta sirve para evaluar la clasificación de un conjunto de datos sin necesidad de una variable de predicción (clasificación no supervisada) [2]. La anchura de la silueta  $s(x_i)$  para el punto  $x_i$  se define como :

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{b(x_i), a(x_i)\}} \quad (9)$$

donde  $x_i$  es un elemento del conglomerado  $\pi_k$ ,  $a(x_i)$  es la distancia media de  $x_i$  a todos los demás elementos del conglomerado  $\pi_k$

$$b(x_i) = \min\{d_l(x_i)\},$$

siendo  $d_l(x_i)$  la distancia media de  $x_i$  a todos los puntos del conglomerado  $\pi_l$  para  $l \neq k$ .

El valor de la anchura de la silueta puede variar entre -1 y 1. Cuanto mayor sea el valor de  $s(x_i)$  mayor es la probabilidad de que esté agrupado en el grupo correcto, y por lo tanto, los valores  $s(x_i)$  negativos tienen más probabilidades de estar agrupados en grupos erróneos [12].

## III. METODOLOGÍA

### III-A. Descripción de los datos

El conjunto de Datos utilizado para este trabajo es de clientes del sector de telecomunicaciones para la predicción de la variable de Abandono (Churn) como se muestra en la tabla I:

Variable	Tipo Dato
Cliente	Id unico
Genero	Catégorico
Tercera Edad	Catégorico
Pareja	Catégorico
Dependientes	Catégorico
Antigüedad	Númerico
Servicio telefónico	Catégorico
Líneas múltiples	Catégorico
Servicio Internet	Catégorico
Seguridad en Línea	Catégorico
Respaldo en Línea	Catégorico
Protección de Dispositivos	Catégorico
Soporte técnico	Catégorico
Televisión por cable	Catégorico
Servicios de Streaming	Catégorico
Contrato	Catégorico
Recibo electrónico	Catégorico
Método de Pago	Catégorico
Cargo mensual	Númerico
Cargos Totales	Númerico
Cliente no Retenido	Catégorico

Tabla I: Variables de entrada

El 26 % de los clientes del conjunto de datos con el que estamos trabajando son clientes no retenidos. para el presente análisis buscamos predecir la categoría de Cliente no retenido.

### III-B. Tratamiento de Datos

La primer parte del proceso es el tratamiento de Datos el cual consta de la limpieza de las variables y transformación para la selección de características, para ello la data fue transformada con el método de transformación Máximo-Mínimo para tener la data en un rango entre 0 y 1, además se aplicó una codificación de las variables categóricas para poder entrenar los modelos.

### III-C. Selección de Variables

Dada la gran cantidad de variables en algunos casos es importante distinguir aquellas variables que son mas relevantes para el análisis por ello en el presente trabajo se aplicó un modelo de selección secuencial de características.

### III-D. Entrenamiento de Modelos

Antes de la ingesta de datos se creó una nueva variable por medio de un algoritmo de clasificación para ver si esta variable permite mejorar las métricas del modelo, para ello se entrenaron 3 de los clasificadores más comunes y se evaluaron la cantidad de subgrupos apropiados y se les asignó una calificación a cada uno por medio de las métricas propuestas para este tipo de modelos [10].

Enseguida se procede a entrenar 5 modelos distintos de clasificación 30 veces cada uno, siendo un total de 150 modelos; Enseguida se aplicará un diseño de experimentos de 5 factores para evaluar si existe alguna diferencia entre la media de la métrica en ellos y algunas pruebas t para confirmar o descartar esta diferencia.

## IV. RESULTADOS EXPERIMENTALES

### IV-A. Selección secuencial de características

De la selección secuencial de características 1 podemos ver que la cantidad óptima de variables oscila entre 9 y 11 variables a elegir, por lo que para el las cuales el rendimiento no aumenta de 0.84 a partir de 10 las 10 variables más importantes las cuales son las siguientes:

- Antigüedad
- Cargos Totales
- Servicio de internet Fibra óptica
- No Servicio de seguridad en línea
- No soporte técnico
- Sin Servicio de televisión
- Contrato mes a mes
- dos contratos al año
- sin recibo impreso
- Método de pago electrónico

### IV-B. Selección Óptima de Subgrupos

Aplicamos el método de Silueta para verificar que realmente la cantidad óptima de subgrupos. Como podemos ver en la figura 2, el valor máximo de subgrupos se alcanza cuando el valor de silueta es de 2 subgrupos.

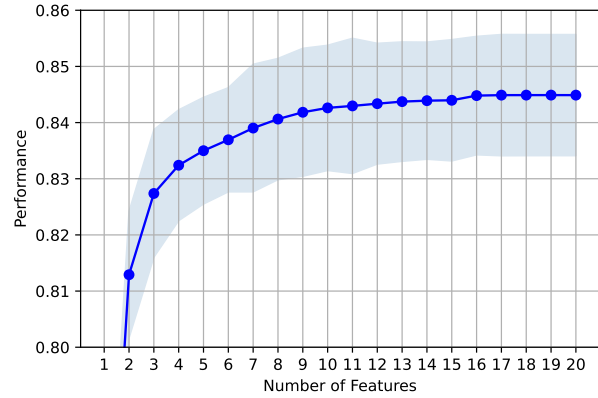


Figura 1: Selección secuencial de características

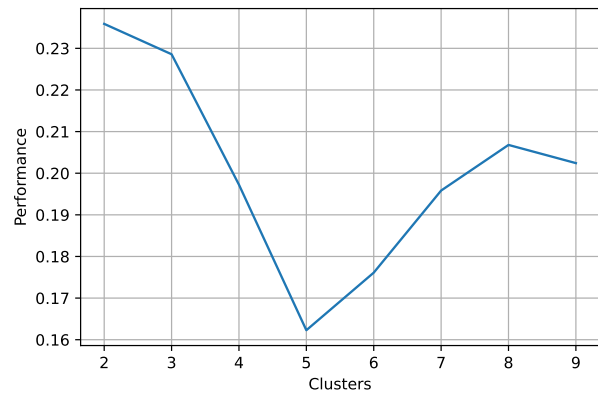


Figura 2: Método de Silueta

### IV-C. Resultados de los modelos no supervisados

Enseguida aplicamos 3 modelos de clasificación no supervisado:

- K-medias
- k-medoides
- Agrupamiento jerárquico

A partir de estos algoritmos de clasificación obtuvimos los siguientes resultados presentados en la tabla II. siendo el mejor modelo el de Agrupamiento jerárquico, el cual tuvo un mejor resultado en la métrica de f1 0.6.

Modelo	accuracy	precision	recall	f1
Ag. Jerarquico	0.72	0.78	0.49	0.60
KMedoides	0.68	0.81	0.45	0.58
kMedias	0.67	0.86	0.43	0.58

Tabla II: Métricas relevantes para la clasificación

La etiqueta generada se almaceno en un un nuevo conjunto para validar si la creación de una variable de grupo mejora las métricas del modelo.

### IV-D. Aplicación de Modelos supervisados

Se entrenaron 150 modelos con el objetivo de realizar un diseño experimental el cuál nos permita encontrar el modelo

más adecuado para este conjunto de datos, por ello se entrenaron los siguientes modelos:

- Regresión Logística (LR)
- LGMB
- Bagging (BC)
- Bosques Aleatorios (RF)
- ADABOOST (ADAB)

obteniendo los resultados reportados en la tabla III.

Modelo	N	$\mu$	$\sigma$	min	25 %	50 %	75 %	max
ADAB	30.00	0.84	0.01	0.82	0.84	0.84	0.85	0.86
BC	30.00	0.77	0.01	0.75	0.77	0.77	0.78	0.79
LGBM	30.00	0.83	0.01	0.80	0.82	0.83	0.84	0.85
LR	30.00	0.84	0.01	0.82	0.84	0.85	0.85	0.86
RF	30.00	0.80	0.01	0.78	0.79	0.80	0.81	0.82

Tabla III: Estadística descriptiva sin variable clase

Enseguida se entrenaron los mismos Modelos con la nueva variable de Clasificación con el objetivo de comparar si existe mejoría en la métrica de AUC, siendo los resultados reportados en la tabla IV.

Modelo	N	$\mu$	$\sigma$	min	25 %	50 %	75 %	max
ADAB	30.00	0.84	0.01	0.82	0.84	0.84	0.85	0.86
BC	30.00	0.79	0.01	0.77	0.78	0.79	0.80	0.82
LGBM	30.00	0.83	0.01	0.82	0.83	0.83	0.84	0.85
LR	30.00	0.84	0.01	0.83	0.84	0.84	0.85	0.86
RF	30.00	0.82	0.01	0.80	0.81	0.82	0.83	0.84

Tabla IV: Estadística descriptiva con variable clase

Al parecer no existe mucha mejora en los algoritmos con mejor AUC, sin embargo los que tuvieron peores resultados en su media incrementaron este indicador al momento de agregar la nueva variable de clasificación.

Enseguida se reportan los resultados de la prueba ANOVA para ver si existe diferencia entre la media de los modelos y algunas pruebas t para comparar aquellos con una media mayor de AUC.

Tipo Anova	SS	DF	MS	F	p-unc
Sin Clase	0.12	4	0.03	299.22	0.00
Con Clase	0.06	4	0.01	136.64	0.00

Tabla V: Anova sin variable clase

vemos que existe diferencia entre las medias de al menos dos variables, por ello aplicamos algunas pruebas t para validar si la diferencia existe entre las variables de mayor AUC del conjunto de datos sin la clasificación reportado en la tabla VI

Modelos	T	dof	alternative	p-val	CI95 %
LR VS ADAB	-0.52	58	two-sided	0.60	[-0.01, 0.0]
LR VS LGBM	4.29	58	two-sided	0.00	[ 0.01, 0.02]
ADAB VS LGBM	5.24	58	two-sided	0.00	[ 0.01, 0.02]

Tabla VI: prueba t de variables con mejor AUC

## V. DISCUSIÓN DE RESULTADOS

Al utilizar el conjunto de datos con la regresión logística obtuvimos una exactitud del 80 %, comparado con un valor de ROC-AUC de 83 %, si utilizamos este modelo para predecir

que clientes retenemos estamos etiquetando bien 8 de cada 10 clientes, esto nos puede ayudar a implementar estrategias para la detección temprana de abandono, sin embargo al agregar la nueva etiqueta y entrenar distintos modelos observamos que los mejores modelos para este conjunto de datos son

- ADA Boost con 0.84 AUC y F1 de 0.59
- Regresión Logística con 0.84 AUC y F1 0.58

los cuales al compararlos con una prueba t, observamos que no existe diferencia entre ambos modelos, lo cual implica que si utilizamos este tipo de modelos en producción obtendremos los mismos resultados. Al agregar una nueva variable de clase no ayuda a mejorar el AUC y F1 en los modelos de mejores resultados (LR, ADAB) sin embargo si mejora la capacidad de predecir de aquellos modelos con un bajo indicador de AUC (RF, LGBM, y BC).

En conclusión, se encontró que algunos algoritmos tienen un mejor rendimiento con este conjunto de datos, y al compararlos se demostró que algunos algoritmos si son superiores que otros para este conjuntos de datos, al agregar una nueva variable de clase, construida a partir de la clasificación natural de los datos mejoró sólo aquellos modelos que en un principio tenían los resultados menos prometedores, por lo que podemos concluir que al entrenar un modelo de clasificación y crear una nueva variable sólo mejorará el modelo si no se tiene un buen resultado, por otro lado al tener un buen resultado no demuestra un mejor resultado al agregarlo al modelo.

Para futuros proyectos se buscará aplicar algoritmos de aprendizaje profundo para mejorar el AUC del modelo y nuevas variables para este conjunto de datos, por otro lado se buscará un conjunto de datos diferente para poder validar si estos modelos se ajustan bien a cualquier modelo de clasificación de retención o sólo para este tipo de casos.

## REFERENCIAS

- [1] A. Siddika, A. Faruque, and A. K. Muhammad Masum, "Comparative analysis of churn predictive models and factor identification in telecom industry," 12 2021, pp. 1–6.
- [2] N. N. Shutaywi, Meshal y Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, 2021. [Online]. Available: <https://www.mdpi.com/1099-4300/23/6/759>
- [3] S. Liu, Rencheng y Ali, S. F. Bilal, Z. Sakhawat, A. Imran, A. Almuhaimeed, and A. Alzahrani, "An intelligent hybrid scheme for customer churn prediction integrating clustering and classification algorithms," *Applied Sciences*, vol. 12, p. 17, 09 2022.
- [4] H. Niu, N. Khozouie, H. Parvin, H. Alinejad-Rokny, A. Beheshti, and M. Mahmoudi, "An ensemble of locally reliable cluster solutions," *Applied Sciences*, vol. 10, p. 1891, 03 2020.
- [5] G. J. y D. Witten, *An Introduction to Statistical Learning with Applications in R*. Los Angeles, CA: Springer, 2021.
- [6] G. Rätsch, T. Onoda, and K.-R. Müller, "Regularizing adaboost," 10 1999.
- [7] R. Aziz, M. F. Baluch, S. Patel, and A. Ganie, "Lgbm: a machine learning approach for ethereum fraud detection," *International Journal of Information Technology*, vol. 14, 01 2022.
- [8] B. Deepa and K. Ramesh, "Epileptic seizure detection using deep learning through min max scaler normalization," *International journal of health sciences*, 05 2022.
- [9] S. Khodabandehlou and M. Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," *Journal of Systems and Information Technology*, vol. 19, pp. 00–00, 08 2017.
- [10] T. Vafeiadis, K. Diamantaras, G. Sarigiannidis, and K. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, 06 2015.

- [11] L. Saha, H. Tripathy, T. Gaber, H. El-Gohary, and E.-S. El-kenawy, "Deep churn prediction method for telecommunication industry," *Sustainability*, vol. 15, p. 4543, 03 2023.
- [12] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0377042787901257>