

# 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 clasificación para identificar aquel algoritmo que mejor clasifique la retención de clientes.

## 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 [?].

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 de acuerdo a la siguiente formula [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 [2].

siendo la función objetivo la siguiente:

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

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 [3].

### 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 siguiente ecuación:

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

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) [4].

### II-C. Métricas

MinMaxScaler es un tipo de normalización que puede escalar todos los valores de un conjunto de datos para que tengan valores de 0 a 1[5], deacuerdo a la ecuación 3.

$$x_s = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

para los modelos de clasificación hay cuatro términos que deben de ser definidos para poder generar las métricas más utilizadas en la literatura [6]:

- 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 4.

$$Exhaustividad = \frac{TP}{TP + FN} \quad (4)$$

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

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

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 [7].

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:

$$F = \frac{2 \times precision \times exhaustividad}{precision + exhaustividad} \quad (6)$$

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 [8]. 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:

$$FPR = \frac{FP}{(FP + TN)} = 1 - TNR \quad (7)$$

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 [1]. La anchura de la silueta  $s(x_i)$  para el punto  $x_i$  se define como :

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{b(x_i), a(x_i)\}} \quad (8)$$

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 [9].

### 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 Maximo-Mínimo para tener la data en un rango entre 0 y 1, además se aplicó una codificación de las variables categóricas para poder ingestarlas en 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 [7].

Enseguida se procede a entrenar un modelo de Regresión Logística y se evaluó si este mejora con y sin el clasificador.

#### 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

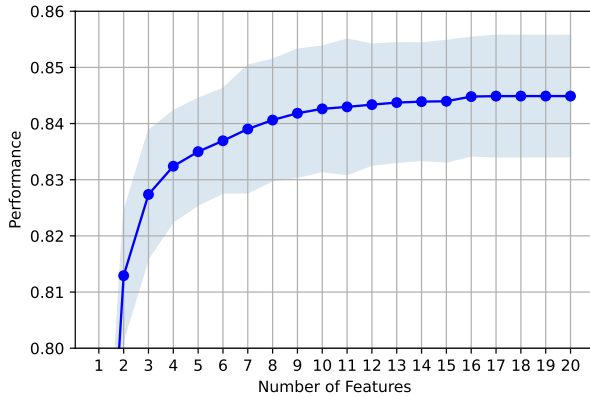


Figura 1: Selección secuencial de características

##### 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.

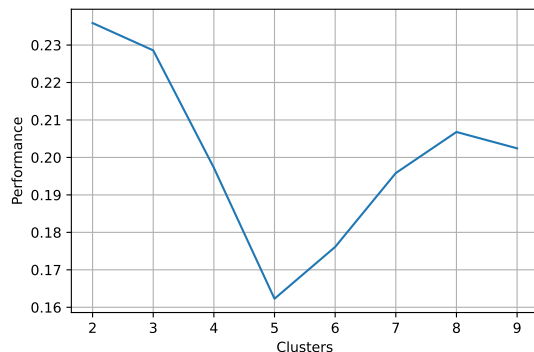


Figura 2: Método de Silueta

##### IV-C. Aplicación de algoritmos de clasificación

Enseguida aplicamos 3 modelos de clasificación no supervisado:

- K-medias
- k-medoides
- Agrupamiento jerárquico

A partir de los 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

##### IV-D. Aplicación de Regresión Logística

Aplicando la regresión logística a los datos transformados con el método de maxMinScaler obtuvimos la curva ROC-AUC en la figura 3, y un valor promedio AUC de 0.83.

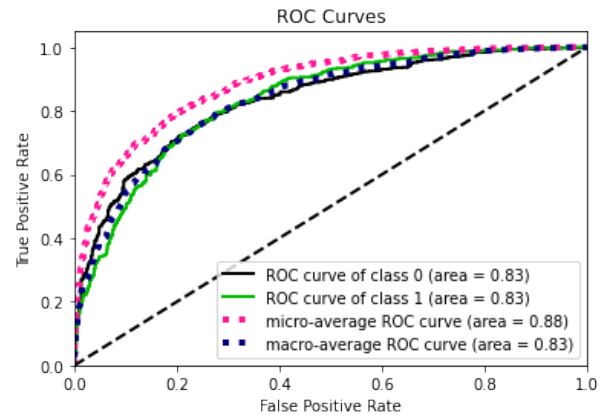


Figura 3: Curva ROC-AUC

#### 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 con el fin de mejorar este indicador se puede utilizar la etiqueta de clasificación como una nueva variable de entrada, por otro lado se podría poner algunos modelos a competir para obtener mejores resultados.

#### REFERENCIAS

- [1] 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>

- [2] 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.
- [3] 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.
- [4] G. J. y D. Witten, *An Introduction to Statistical Learning with Applications in R*. Los Angeles, CA: Springer, 2021.
- [5] B. Deepa and K. Ramesh, "Epileptic seizure detection using deep learning through min max scaler normalization," *International journal of health sciences*, 05 2022.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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>