

**MODELO DE APRENDIZAJE AUTOMÁTICO BASADO EN REDES LSTM
OPTIMIZADAS CON ALGORITMOS METAHEURÍSTICOS PARA
PREDECIR LA DESERCIÓN DE TARJETAHABIENTES**

Por:

Diana Marcela Correa Jaramillo

Trabajo de grado optar al grado de
Magíster en Ciencias de los Datos y Analítica

Director:

Jose Lisandro Aguilar Castro

Mes del 2023

Medellín

Resumen

Uno de los principales retos en la industria, es entender el comportamiento de sus clientes y anticiparse a sus posibles acciones con el fin de mantener relaciones a largo plazo. Por lo que disponer de herramientas que permitan a las compañías una acertada toma de decisiones es una tarea que ha venido tomando mayor relevancia en la actualidad. Dicho panorama no es hecho aislado para las entidades financieras, las cuales, debido a la amplia oferta de productos similares ofrecidos por otras compañías deben constantemente definir y estructurar nuevas estrategias que permitan la retención y fidelización del cliente.

Este trabajo parte una exploración sobre diferentes técnicas de aprendizaje automático que permiten anticiparse a la identificación oportuna de los clientes que poseen al menos una tarjeta de crédito y tienen riesgo de cancelar el producto. En ésta, se identifica algunas de las técnicas mayormente empleadas para identificar la deserción del cliente en los diferentes nichos empresariales como lo son el uso del SVM, Arboles de decisión, Bosques aleatorios, Redes neuronales, entre otras; dichas técnicas según el caso de estudio han presentado un buen desempeño en el momento de identificar los clientes con riesgo de fuga, pero a su vez no son tan precisas cuando se consideran las dependencias y patrones en largos periodos de tiempo. Es por ello, que con este se plantea el uso de redes neuronales LSTM, las cuales tiene como propiedad recordar patrones a largo plazo, adicionalmente se formula un enfoque híbrido mediante una la red LSTM optimizada con un algoritmo genético. Los resultados experimentales obtenidos empelando un enfoque temporal del problema mediante redes LSTM y un enfoque tradicional mediante el uso de técnicas como SVM y Bosques aleatorios son comparados al finalizar.

Tabla de contenido

Resumen	2
Tabla de contenido.....	3
Listado de figuras.....	4
Listado de tablas	4
Listado de abreviaturas.....	5
Introducción	6
1. Estado del arte y Marco teórico	7
1.1. Deserción de clientes en diferentes sectores empresariales	7
1.2. Métodos tradicionales de aprendizaje automático para la predicción en la deserción del tarjetahabiente.....	8
1.2.1. Preprocesamiento de datos	9
1.2.2. Modelado	10
1.2.2.1. Regresión logística.....	10
1.2.2.2. Árboles de decisión	10
1.2.2.3. Calificador Naïve Bayes.....	10
1.2.2.4. K-vecinos más cercanos (K-NN)	11
1.2.2.5. Máquinas de vectores de soporte (SVM).....	11
1.2.3. Evaluación modelo	13
1.2.3.1. Métricas escalares.....	13
1.2.3.2. Métricas basadas en gráficos.....	15
1.3. Métodos basados en series temporales para la predicción en la deserción del tarjetahabiente.....	15
1.3.1. Red neuronal LSTM (Long Short-Term Memory)	15
1.3.1.1. Ventajas y limitaciones en la implementación de una red neuronal LSTM	18
1.3.1.2. Implementaciones híbridas para la optimización de los hiperparámetros de una red neuronal LSTM 18	
1.3.1.2.1. Algoritmos heurísticos para la optimización	18
1.3.1.2.1.1. Algoritmo genético.....	18
1.3.1.2.1.2. Algoritmo PCO.....	19
2. Metodología	20
3. Resultados y análisis	21
3.1. Información básica sobre los datos empleados	21
3.2. Análisis exploratorio de los datos	21
3.3. Aplicación de los modelos y análisis de los resultados	21
4. Conclusiones.....	21
5. Recomendaciones.....	21

Listado de figuras

Ilustración 1: Ejemplo K-vecinos más cercanos (editado de [18]).....	11
Ilustración 2: Descripción bosques aleatorios (editado de [18]).....	12
Ilustración 3: Esquema neuronal artificial (editado de [32] ,[32])	13
Ilustración 4: Arquitectura de las células de memoria de una LSTM (editado de [23], [24], [34]).....	16
Ilustración 5: Esquema funcionamiento algoritmo genético (editado de [15], [37], [38], [46])	19
Ilustración 6: Ciclo de vida de minería de datos (tomado de [29])	20
Ilustración 7: Flujo aprendizaje automático para la predicción y el análisis de abandono, (editado de [14], [16], [23])	21

Listado de tablas

Tabla 1: Literatura relacionada sobre la tasa de abandono de clientes.....	8
Tabla 2: Métodos de muestreo	9

Listado de abreviaturas

- **AUC** Area under the ROC curve (área bajo la curva ROC)
- **DT** *Decision Tree* (árbol de decisión).
- **K-NN** K-nearest neighbors (k-vecinos más cercanos)
- **LSTM** *Long Short-Term Memory*.
- **PSO** *Particle Swarm Optimization* (Optimización de Enjambre de partículas)
- **ROC** Receiver operating characteristics (ROC)
- **SVM** *Support Vector Machine* (máquina de vectores de soporte)
- **RF** *Random Forest* (bosque aleatorio)
- **GA** Genetic Algorithm

Introducción

En la actualidad, las entidades financieras disponen de un gran número de clientes, los cuales son cada vez más conscientes de la calidad del servicio y productos que reciben, dada la amplia oferta de productos similares por parte de otros bancos. La creciente competencia entre las entidades financieras ha dado pie a la necesidad de definir y estructurar mejores estrategias que permitan la retención del cliente, pero a su vez, que contribuyan con un aumento en la calidad del servicio prestado [1],[2],[3].

Uno de los pilares fundamentales de la banca moderna es el manejo del cliente [4], lo que implica entender su comportamiento, siendo este es uno de los principales retos a ser afrontados en esta industria, ya que perder un cliente significa una reducción de potenciales ganancias para la entidad. Es por ello, que constantemente se están generando nuevas herramientas que permiten a partir de la información disponible, entender y pronosticar las posibles acciones del cliente, conocer sus niveles de satisfacción, sus necesidades de cara al servicio prestado, además de identificar aquellos que poseen un potencial riesgo de deserción. Esto le permite a la compañía gestionar proactivamente dicho riesgo, ya sea generando estrategias para fortalecer su relación comercial, u ofreciendo ofertas y productos mucho más atractivos que la competencia [1], [4], [5]. Para ello, el sector financiero, se ha apoyado de técnicas de aprendizaje automático que le permiten mejorar su operación reduciendo costos, optimizando el análisis del riesgo y generando ofertas personalizadas [6].

Entre los diferentes productos ofrecidos por las entidades financieras, se encuentran las tarjetas de crédito, cuyo uso se ha extendido durante las últimas décadas, de tal forma que en 2013 se estimaba que aproximadamente la mitad de la población adulta en países desarrollados, poseían al menos una tarjeta de crédito [7]. Dicha situación no ha sido ajena en Colombia, en donde para el año 2022, la proyección poblacional era de aproximadamente 51.6 millones de habitantes [8], al tiempo que para noviembre de 2022, según Asobancaria, el número de tarjetas de crédito vigentes emitidas, ya sea por cooperativas, compañías de financiamiento comercial o establecimientos bancarios, era de 16.1 millones de tarjetas [9]. Debido al relevante número de tarjetas activas respecto a la población, y dado que este producto financiero es de fácil acceso y dispone de una amplia oferta por múltiples compañías, es fundamental estudiar el comportamiento de los tarjetahabientes. Por lo que, para una entidad financiera, disponer de un modelo predictivo que permita identificar oportunamente aquellos tarjetahabientes con riesgo de deserción es de vital importancia, ya que fidelizar un cliente resulta más económico que atraer uno nuevo [10]. Además, aporta información para generar estrategias más efectivas para la retención del cliente y/o incentivar el uso del producto, permitiendo de esta manera ser más competitiva respecto a otras compañías que prestan el mismo servicio [2],[6].

MODELO DE APRENDIZAJE AUTOMÁTICO BASADO EN REDES LSTM OPTIMIZADAS CON ALGORITMOS METAHEURÍSTICOS PARA PREDECIR LA DESERCIÓN DE TARJETAHABIENTES

1. Estado del arte y Marco teórico

1.1. Deserción de clientes en diferentes sectores empresariales

En la actualidad, los clientes son considerados uno de los activos más valiosos en las organizaciones, lo que ha generado un creciente interés en el desarrollo de herramientas que apoyen en la retención y predicción de posibles fugas de estos, extendiéndose a diferentes industrias, tales como las telecomunicaciones, los seguros, el sector financiero, el comercio minorista y otros [11],[12],[13],[14],[15],[16]. Dichas herramientas deben permitir identificar las causas por las que se presenta la deserción o rotación del cliente, partiendo de los grandes volúmenes de información disponibles [17], con miras a la generación de estrategias que permitan la acertada toma de decisiones [18].

Hoy en día, los servicios prestados por un proveedor en particular no difieren en gran proporción de los prestados por otro dentro de su mismo nicho. Un ejemplo de esto se ve en los servicios prestados por una entidad financiera, lo que ha generado un ambiente cada vez más competitivo entre ellas, llevándolos a reevaluar las políticas existentes en el manejo de la deserción y el dimensionamiento de los beneficios reales para la preservación y fidelización del cliente. Por ejemplo, un aumento del 5 % en la tasa de retención implica un aumento de hasta un 85 % en sus beneficios [18],[11]. Luego, para ser una empresa líder, no sólo es necesario estar orientado al cliente, sino además centrarse en la creación de relaciones a largo plazo [19]. Para ello, se ha generado una creciente demanda en la predicción oportuna de la rotación de los clientes, en donde ha sido necesario el uso de técnicas estadísticas y de minería de datos para la generación de modelos de predicción, tales como redes neuronales artificiales, árboles de decisión, bosques aleatorios, máquinas de vectores de soporte (SVM, por sus siglas en inglés), regresiones logísticas, entre otras [11], [16].

A continuación, se detallan algunos ejemplos de estudios realizados para la predicción de la deserción de clientes empleando diversas técnicas en diferentes nichos empresariales (ver

Tabla 1), en donde los resultados obtenidos difieren entre sí en efectividad de acuerdo con las consideraciones realizadas, el tratamiento y manejo de los datos.

Tabla 1: Literatura relacionada sobre la tasa de abandono de clientes

Nicho	Autor	Métodos de predicción
Telecomunicaciones	Keramati y otros [20]	Arboles de decisión, Redes neuronales, K-vecinos más cercanos, SVM
	Xia y Jin [21]	SVM, Redes neuronales, Arboles de decisión, Regresión logística, Clasificador Naïve Bayes
	Sharma y Kumar [22]	Redes neuronales
	Huang y otros [23]	Regresión logística, Arboles de decisión, Clasificador Naïve Bayes, Clasificador Lineal, Redes neuronales, SVM, Algoritmo de minería de datos evolutivos
	Amin y otros [15]	Clasificador Naïve Bayes con un algoritmo genético
Comercio electrónico	Yu y otros[24]	ESVM (Extended Support Vector Machine), Redes neuronales, Arboles de decisión, SVM
Financiero	Nie y otros [19]	Regresión logística, Arboles de decisión
	Lin y otros [25]	Modelo híbrido basado en un enfoque de conjunto aproximado y grafos para realizar la predicción
	Kumar y Ravi [1]	Regresión logística, Arboles de decisión, Bosques aleatorios, Perceptrón multicapa, Función de base radial, SVM
	Miao y Wang [18]	Bosques aleatorios, Regresión logística, K-vecinos más cercanos
Servicios de suscripción	Coussement y otros [26]	SVM, Regresión logística, Bosques aleatorios

1.2. Métodos tradicionales de aprendizaje automático para la predicción en la deserción del tarjetahabiente

La predicción en la rotación del cliente presenta un problema de desequilibrio de clases [14],[27], dado a que para un periodo de tiempo, la cantidad de tarjetahabientes que continúan con el producto es mayor a la cantidad de clientes que cancelan su tarjeta de crédito, lo que conlleva a problemas en el momento de realizar una predicción acertada si no se realiza un adecuado

tratamiento de los datos antes de entrenar el modelo, a raíz de esto se propone el siguiente enfoque de trabajo [27].

1.2.1. Preprocesamiento de datos

Un adecuado preprocesamiento de los datos influye en el rendimiento de la implementación realizada [15], por lo que cuando se dispone de un problema de desbalanceo clases, es de vital importancia emplear técnicas que permitan cambiar la distribución de los datos para evitar que el modelo se encuentre sesgado hacia la clase mayoritaria (no deserción) [14],[16],[27], dichas técnicas pueden agruparse de la siguiente manera:

- **Cambio de distribución:** Consiste en cambiar la distribución de los datos con el fin de obtener una base más equilibrada. En donde, para problemas con desequilibrio extremo, se tiene un número de ejemplos normales mucho mayor que el número de ejemplos raros, un ejemplo de esto son los clientes que desertan; en estos casos se recomienda un equilibrio entre clases de 2:1 o 3:1 (mayoría: minoría) [27],[28]. A continuación describen algunos enfoques para cambiar la distribución de los datos.

Tabla 2: Métodos de muestreo

Muestreo	Descripción
Sobremuestreo u <i>Oversampling</i>	Consiste en duplicar los registros de la clase minoritaria o generar registros sintéticos nuevos a partir de los disponibles. Entre las técnicas más utilizadas se tienen la técnica de sobremuestreo de minorías sintéticas (SMOTE) y el muestreo sintético adaptativo (ADASYN) [14],[16],[29].
Submuestreo o <i>Undersampling</i>	Consiste en eliminar registros de la clase mayoritaria. Una de las técnicas de submuestreo es la regla de limpieza de vecindario [14],[16].
Hibrido	Al emplear un método híbrido es posible obtener una mejor separación de las clases. En éste se han propuesto varias combinaciones de muestreo y submuestreo tales como [14], [16]: <ul style="list-style-type: none"> • SMOTE y <i>Random Undersampling</i>, en el cual se aumenta el tamaño de la clase minoritaria tomando aquellas instancias similares y sintetizando nuevas [14]. • SMOTE con <i>Tomek Links</i>. En esta mediante SMOTE se sobremuestra la clase minoritaria y luego mediante el submuestreo de enlaces de Tomek se produce un conjunto de datos equilibrado [14].

- **Ponderar el espacio de los datos:** Permite modificar la distribución del conjunto de datos empleando información asociada los costos de clasificar erróneamente un dato [27].

1.2.2. Modelado

Desde diferentes investigaciones, se ha abarcado la temática de la predicción en la rotación del cliente [30]. En esta sección se hace un recorrido por diferentes técnicas descritas en

Tabla 1.

1.2.2.1. Regresión logística

Mediante una regresión logística se modela la probabilidad de deserción empleando una función lineal. El modelo se puede escribir de la siguiente manera [14],[23]:

$$prob(y = 1) = \frac{e^{\beta_0 + \sum_{k=1}^K \beta_k x_k}}{1 + e^{\beta_0 + \sum_{k=1}^K \beta_k x_k}} \quad (1)$$

Donde:

- Y es la variable dependiente que representa el evento (rotación del cliente).
- x_1, x_2, \dots, x_k son las entradas
- $\beta_1, \beta_2, \dots, \beta_k$ son los coeficientes de regresión

1.2.2.2. Árboles de decisión

Los árboles de decisión inicialmente buscan el atributo con el que se obtiene la mejor ganancia de información en el nodo raíz de manera recursiva, para ello se divide el árbol en subárboles y a cada uno de estos se le busca el mejor ajuste hasta alcanzar una regla de parada [14],[23].

1.2.2.3. Calificador Naïve Bayes

Calcula la probabilidad de que una muestra de entrada dada pertenezca a una clase determinada. En donde dada una muestra $X = \{x_1, x_2, \dots, x_n\}$, la probabilidad para que la clase y_j pertenezca a una clase, se encuentra dada mediante la siguiente ecuación [15], [23]:

$$p(y_i|X) = p(X|y_j)p(y_j) = p(x_1, x_2, \dots, x_n|y_i)p(y_i) \quad (2)$$

Donde:

- $p(y_i)$ es la probabilidad a priori de y_i
- $p(X|y_j) = \prod_{i=1}^n p(x_i|y_j)$ (3)

1.2.2.4. K-vecinos más cercanos (K-NN)

El algoritmo K-NN asigna a una instancia x_i la etiqueta que corresponde a la etiqueta mayoritaria de las k muestras tomadas desde el entrenamiento, sin embargo, presenta problemas significativos cuando se emplea en problemas de deserción [14]. En la Ilustración 1 se da un ejemplo de cómo un dato recién ingresado a una muestra (circulo azul) debe clasificarse entre triángulos o cuadrados, en donde si se toma un $K = 3$ (área dentro de línea sólida), se clasificaría como un triángulo, mientras que si se considera un $K = 5$ (área dentro de línea punteada) quedaría clasificado como un cuadrado [18].

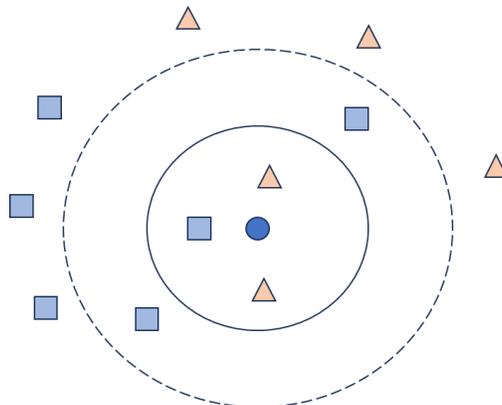


Ilustración 1: Ejemplo K-vecinos más cercanos (editado de [18])

1.2.2.5. Máquinas de vectores de soporte (SVM)

Las máquinas de vectores de soporte o SVM (por sus siglas en inglés) es un modelo de aprendizaje automático basado en kernel empleado en tareas de clasificación y regresión. En este se construye un hiperplano de separación óptimo entre las clases linealmente separables, el cual solo considera los vectores de soporte [14], [23], [26], [31], lo que lo convierte en una alternativa interesante para conjuntos de datos moderadamente desequilibrados, aunque tiene un desempeño deficiente cuando la distribución de clases es demasiado sesgada [14].

1.2.2.6. Bosques aleatorios (RF)

Los bosques aleatorios emplean un subconjunto de m predictores, los cuales son elegidos de manera aleatoria para generar cada árbol en una muestra con los datos de entrenamiento. En donde para predecir la etiqueta a ser asignada a la clase, se toma la clase más popular [26].

Es un algoritmo de fácil implementación, dado a que sólo se debe considerar el número de predictores a emplear, el cual se sugiere sea igual a la raíz cuadrada del número total de variables y un gran número de árboles [26]. En la **¡Error! No se encuentra el origen de la referencia.** se muestra un ejemplo simple de un bosque aleatorio.

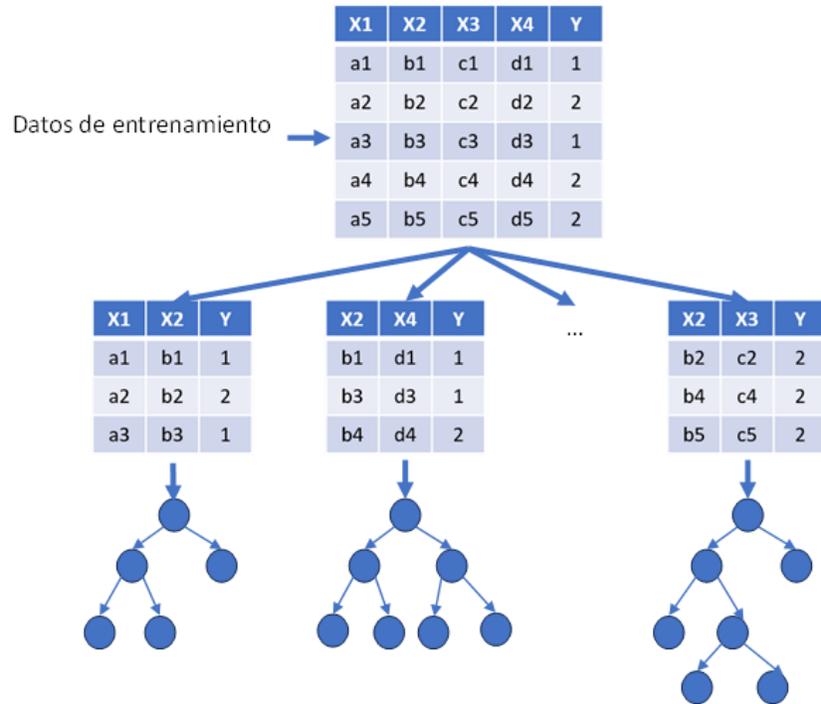


Ilustración 2: Descripción bosques aleatorios (editado de [18])

1.2.2.7. Redes neuronales artificiales

Las redes neuronales son una abstracción del funcionamiento del cerebro humano [31], [32], éstas son una red compleja compuesta por un conjunto de nodos llamados neuronas que realizan operaciones matemáticas definidas [32], [33]. Las redes neuronales, toman los datos de entrada y con éstos inicia una etapa de aprendizaje la cual puede ser supervisada, no supervisada o por refuerzo, dónde se obtiene diferentes características de los datos para realizar la predicción [34]. Existen muchos tipos de redes neuronales que varían de acuerdo con el uso y complejidad, en donde se tienen redes como el perceptrón, las redes neuronales recurrentes, las redes convolucionales, entre otras [32], [34].

En la Ilustración 3, se puede ver el esquema de una red neuronal artificial, en donde ésta toma la información que llega mediante las entradas junto con el sesgo, se suman y son procesadas mediante una función de transferencia para finalmente obtener una salida [32], [34].

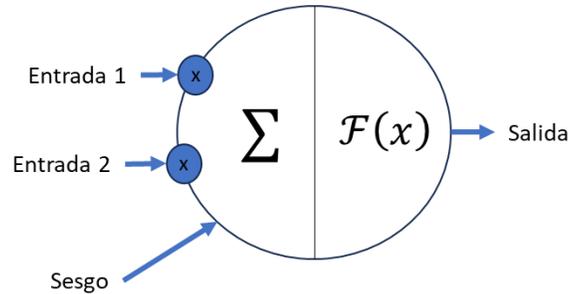


Ilustración 3: Esquema neuronal artificial (editado de [32] ,[32])

La descripción matemática de las neuronas artificiales se describe a continuación [34]:

$$y(x) = F(\sum_{i=0}^m w_i(k) x_i(k) + b) \quad (4)$$

Donde:

- $x_i(k)$ es el valor de la entrada en el tiempo k.
- $w_i(k)$ es el valor del peso en el tiempo k.
- b es el sesgo.
- F es la función de transferencia
- $y(x)$ es el valor de la salida en el tiempo k

Cuando se tienen problemas para los cuales debe aprender y almacenar información a muy largo tiempo, las redes neuronales recurrentes estándar carecen de dependencia a largo plazo, por lo que es necesario emplear redes LSTM, las cuales son redes neuronales recurrentes más avanzadas que tienen la capacidad de aprender dependencias a largo plazo [34]. Para ampliar información sobre redes LSTM, remítase a: Red neuronal LSTM (Long Short-Term Memory)

1.2.3. Evaluación modelo

En la evaluación del desempeño cuando se tienen problemas de datos desequilibrados se presentan desafíos cuando se busca predecir la clase que se encuentra poco representada dentro de la muestra de datos disponible, ya que los criterios de evaluación tradicionales tienden a centrar la evaluación en los casos que más se presentan, por lo que emplear este tipo de métricas pueden generar conclusiones erróneas, es por ello que se debe recurrir a otras técnicas de evaluación, como las que se muestran a continuación [27]:

1.2.3.1. Métricas escalares

$$\text{true positive rate (recall or sensitivity): } TP_{rate} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{true negative rate (specificity): } TN_{rate} = \frac{TN}{TN+FP} \quad (6)$$

$$\text{false positive rate: } FP_{rate} = \frac{FP}{TN+FP} \quad (7)$$

$$\text{false negative rate : } FN_{rate} = \frac{FN}{TP+FN} \quad (8)$$

$$\text{positive predictive value (precision): } PP_{value} = \frac{TP}{TP+FP} \quad (9)$$

$$\text{negative predictive value : } NP_{value} = \frac{TN}{TN+FN} \quad (10)$$

$$F_{\beta} = \frac{(1+\beta)^2 * recall * precision}{\beta^2 * precision + recall} \quad (11)$$

Donde β es un coeficiente establecido para ajustar la importancia relativa entre el *recall* y la precisión.

Nota: Usualmente, cuando se está trabajando con datos desequilibrados, para el cálculo de F_{β} , se da la misma importancia al *recall* y la precisión, es decir $\beta = 1$ [27].

$$G - Mean = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}} = \sqrt{\text{sensitivity} \times \text{specificity}} \quad (12)$$

G-Mean calcula la media geométrica de las precisiones de las dos clases. Fue desarrollada específicamente para evaluar el rendimiento en dominios desequilibrados [27].

1.2.3.2. Métricas basadas en gráficos

Dos métricas basadas en gráficos empleadas cuando se tienen un problema con datos desbalanceados son la curva ROC y el área bajo la curva ROC (AUC), donde la primera permite la visualización de la compensación relativa entre TP_{rate} y FP_{rate} , mientras que el AUC permite la evaluación del mejor modelo en promedio [27], [28], en donde valores de AUC más altos indica que el clasificador tiene un mejor rendimiento [28].

Nota: Dentro del contexto de la deserción, las métricas recomendadas son la F1-score y el AUC. En donde la primera entrega un valor estimado entre el *recall* y la precisión mientras que la segunda proporciona la tasa de verdaderos positivos en función de la tasa de falsos positivos [16].

1.3. Métodos basados en series temporales para la predicción en la deserción del tarjetahabiente

En el apartado anterior, se mostraron algunos de los modelos empleados para predecir el abandono del cliente partiendo de su comportamiento estático, en la actualidad, se ha venido explorando un enfoque dinámico empleando el uso de redes neuronales a largo plazo (LSTM), con las que se busca identificar en el comportamiento del cliente a lo largo del tiempo para predecir un potencial riesgo de fuga éste [35], [36].

1.3.1. Red neuronal LSTM (Long Short-Term Memory)

Las redes neuronales recurrentes (RNN) tienen la propiedad conservar las variables de estado a lo largo del tiempo [36], [37], lo que permite el aprendizaje secuencial. Pero estas solo pueden manejar dependencias a corto plazo, por lo cual se presenta el problema del gradiente de fuga [34], [38]. Las redes LSTM pertenecen a las RNN, y han mostrado resultados prometedores cuando son implementadas en tareas con series temporales. La arquitectura de este tipo de redes tiene tres tipos de capas [34], [36], [38], [39], [40].

- **Capa de entrada:** La cual tiene un número de neuronas igual al número de variables de entrada.
- **Capas ocultas simples o múltiples:** Las cuales están formadas por células de memoria.
- **Capa de salida:** Que dispone de un número neuronas igual al número de variables de salida.

Las redes LSTM son superiores a la RNN estándar debido a que estas disponen de una célula de memoria (celda de estado LSTM) que puede retener información a lo largo del tiempo. La estructura de la célula de memoria es definida por tres tipos de puertas, las cuales son [34], [38], [39], [40]:

- **Puerta de olvido (f_t):** Toma las entradas y decide qué información se va a eliminar de la memoria.
- **Puerta de entrada (i_t):** Decide qué información será añadida a la memoria.
- **Puerta de salida (o_t):** Decide la salida del bloque de memoria.

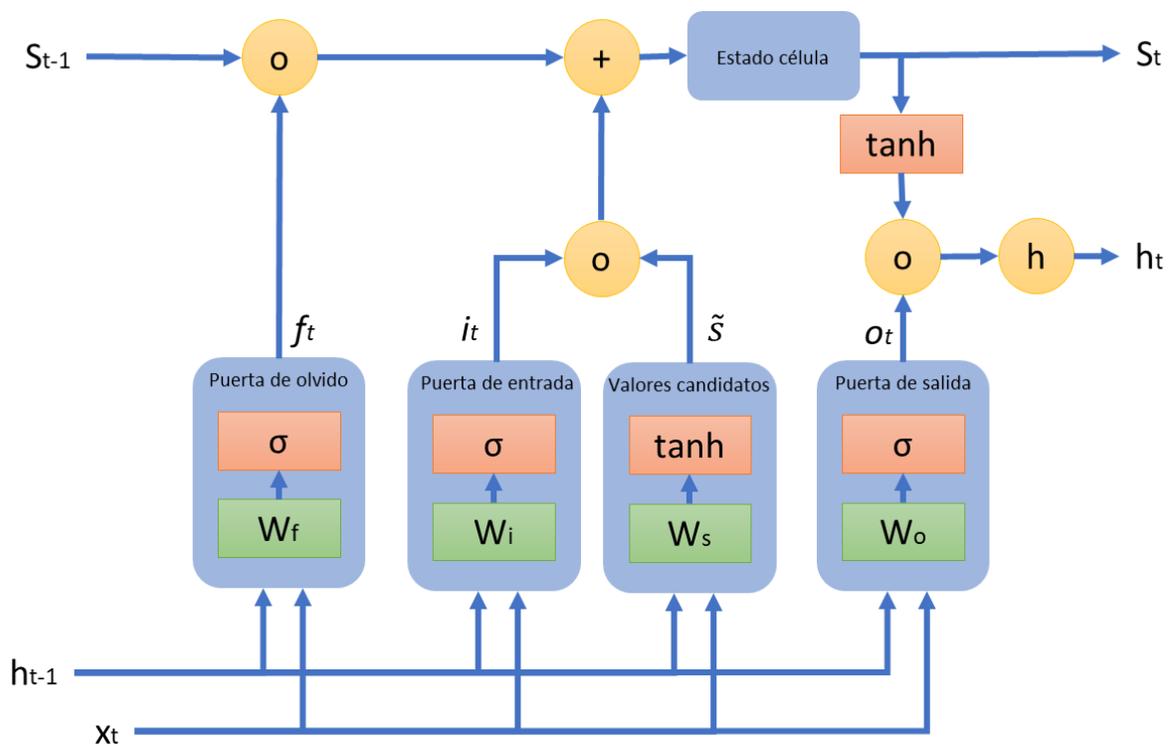


Ilustración 4: Arquitectura de las células de memoria de una LSTM (editado de [23], [24], [34])

A continuación, se detallan las ecuaciones que describen el comportamiento de las celdas de estado de las redes LSTM, considerando la siguiente notación [39]:

- S_{t-1} corresponden a los vectores de estado de la celda en el tiempo anterior.
- S_t , \tilde{S}_t corresponden a los vectores de estado de la celda y los valores candidatos respectivamente.
- h_t es el vector de salida de la capa LSTM.
- h_{t-1} es el vector de salida de la capa LSTM en el tiempo anterior.

- X_t es el vector de entradas en el tiempo t .
- $W_{f,x}, W_{f,h}, W_{\tilde{s},x}, W_{\tilde{s},h}, W_{i,x}, W_{i,h}, W_{o,x}, W_{o,h}$ son las matrices de peso.
- $b_f, b_{\tilde{s}}, b_i, b_o$ son los vectores de sesgo.
- f_t, i_t, o_t son los vectores de activación para las compuestas de olvido, entrada y salida respectivamente.

El comportamiento de una celda de estado LSTM es definido como [34], [38], [39]:

- **Paso 1:** se ejecuta la capa del olvido, una capa sigmoidea que es una función de activación no lineal que toma como entradas el elemento de la secuencia de entrada X_t , la salida del paso de tiempo anterior h_{t-1} , el sesgo b_f , para calcular el vector de activación f_t para cada uno de los valores del estado de la celda S_{t-1} normalizados entre 0 y 1. Esta capa indica qué se olvida completamente. A continuación, se detalla la expresión para el vector de valores de activación.

$$f_t = \text{sigmoide} (W_{f,x} X_t + W_{f,h} h_{t-1} + b_f) \quad (13)$$

- **Paso 2:** En este paso se decide la información que será añadida al estado de la célula de memoria S_t . Este paso se encuentra compuesto por dos partes, la primera de ellas calcula los valores candidatos \tilde{S}_t , y la segunda parte dispone de una capa de activación llamada capa de puerta de entrada i_t .

$$\tilde{S}_t = \tanh (W_{\tilde{s},x} X_t + W_{\tilde{s},h} h_{t-1} + b_{\tilde{s}}) \quad (14)$$

$$i_t = \text{sigmoide} (W_{i,x} X_t + W_{i,h} h_{t-1} + b_i) \quad (15)$$

- **Paso 3:** Se calcula los nuevos estados para la celda S_t .

$$s_t = f_t \circ S_{t-1} + i_t \circ \tilde{S}_t \quad (16)$$

- **Paso 4:** Se calcula los nuevos estados para la salida de la celda.

$$O_t = \text{sigmoide} (W_{o,x} X_t + W_{o,h} h_{t-1} + b_o) \quad (17)$$

$$h_t = O_t \circ \tanh(s_t) \tag{18}$$

Tal y como se muestra en la Ilustración 4, las variables de entrada se insertan en las puertas de entrada de la celda de estado LSTM. Estas son procesadas empleando las ecuaciones descritas de la 12 a 17, y luego de finalizado el proceso, se genera la secuencia de salida o pronóstico de la red [38], [39].

1.3.1.1. Ventajas y limitaciones en la implementación de una red neuronal LSTM

Cuando se disponen de datos secuenciales, los cuales a su vez poseen una brecha de tiempo entre los mismos, el uso de redes LSTM generan un resultado superior, debido a que no solo poseen la capacidad de actualizar su estado partiendo de los estados pasados de esta y los datos de entrada actuales del proceso, sino que, permiten la conservación de las variables de estado a lo largo del tiempo, lo que contribuye a disponer de una potente capacidad de aprendizaje [41]. Este tipo de implementaciones requiere de una adecuada selección de los hiperparámetros de la red para obtener mejores resultados, lo que implica un aumento en la complejidad.

1.3.1.2. Implementaciones híbridas para la optimización de los hiperparámetros de una red neuronal LSTM

La optimización en la selección de los hiperparámetros de la red LSTM dentro de este alcance, se acota al problema en la determinación del tamaño de la ventana temporal empleada en el problema de clasificación abordado. Para ello, se explora un enfoque híbrido mediante el uso algoritmos heurísticos para la optimización.

1.3.1.2.1. Algoritmos heurísticos para la optimización

En la actualidad existen una gran cantidad de problemas de optimización, muchos de los cuales debido a su complejidad no se pueden resolver en un tiempo polinómico utilizando métodos exactos, por lo que es necesario implementar nuevas técnicas que permitan obtener una solución aceptable en un tiempo razonable. Dichas técnicas son conocidas como heurísticas, y poseen un esquema de búsqueda el cual se basa en la exploración o diversificación y la explotación o intensificación, lo que las hace muy robustas [42]. Pero a pesar de esto, no todos los problemas de optimización se pueden resolver con la misma técnica, por lo que se han desarrollado una gran cantidad de heurísticas, las cuales pueden ser inspiradas en la naturaleza (metaheurísticas) o no [42].

1.3.1.2.1.1. Algoritmo genético

Es una subclase de los algoritmos evolutivos, los cuales, son ampliamente empleados en procesos de iteración en los que emplean una población de soluciones y operaciones como selección, cruce, mutación y reproducción para obtener soluciones cada vez mejores (cromosomas) [15], [34], [43], [44], [45].

El algoritmo genético funciona de la siguiente manera:

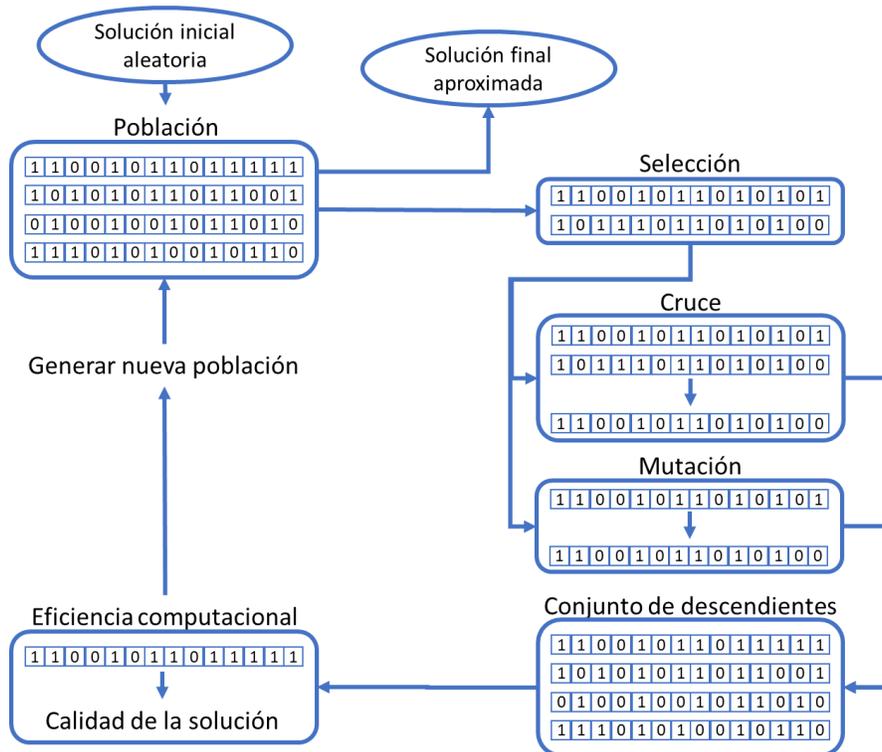


Ilustración 5: Esquema funcionamiento algoritmo genético (editado de [15], [37], [38], [46])

1.3.1.2.1.2. Algoritmo PCO

Es una subclase de los algoritmos de enjambre, en el cual se imita el comportamiento colectivo como el de los pájaros (agentes) [42].

2. Metodología

La metodología empleada es CRISP-DM (*Cross Industry Standard Process for Data Mining*), la cual fue elegida debido a su flexibilidad y fácil personalización de acuerdo con las necesidades del caso. Para ello, se ha tomado como base la guía descrita por IBM [47] (Ver **¡Error! No se encuentra el origen de la referencia.**), considerando las etapas que van desde la comprensión del negocio hasta la evaluación de modelo predictivo. El despliegue no está incluido dentro del alcance, debido a que para el mismo se deben cumplir unos lineamientos, consideraciones y políticas que son definidos por la entidad financiera que suministra los datos.

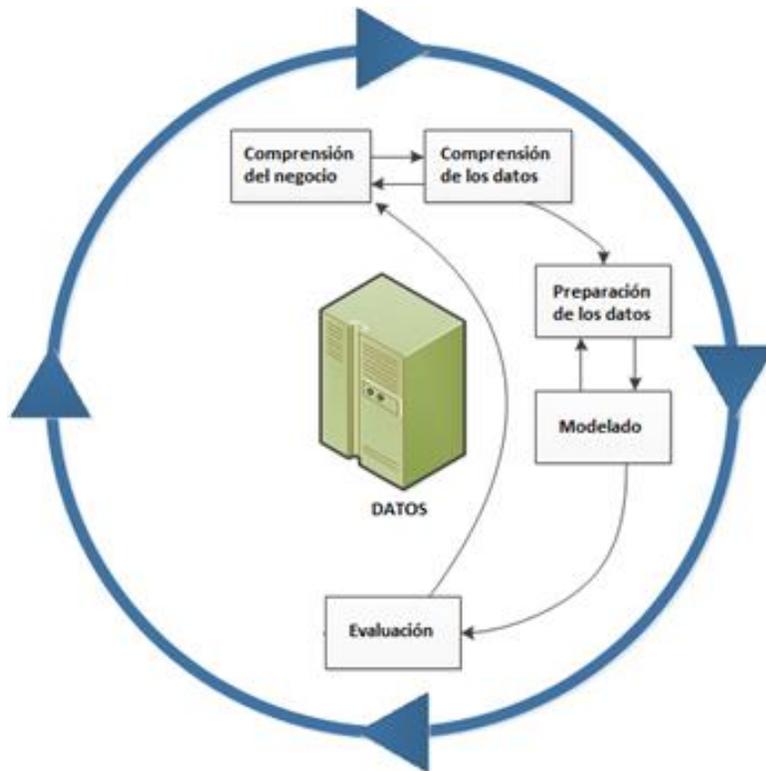


Ilustración 6: Ciclo de vida de minería de datos (tomado de [29])

El flujo de trabajo empleado para la predicción en la rotación de clientes dentro de las etapas contempladas entre la preparación de los datos y la evaluación se resume en la Ilustración 7.

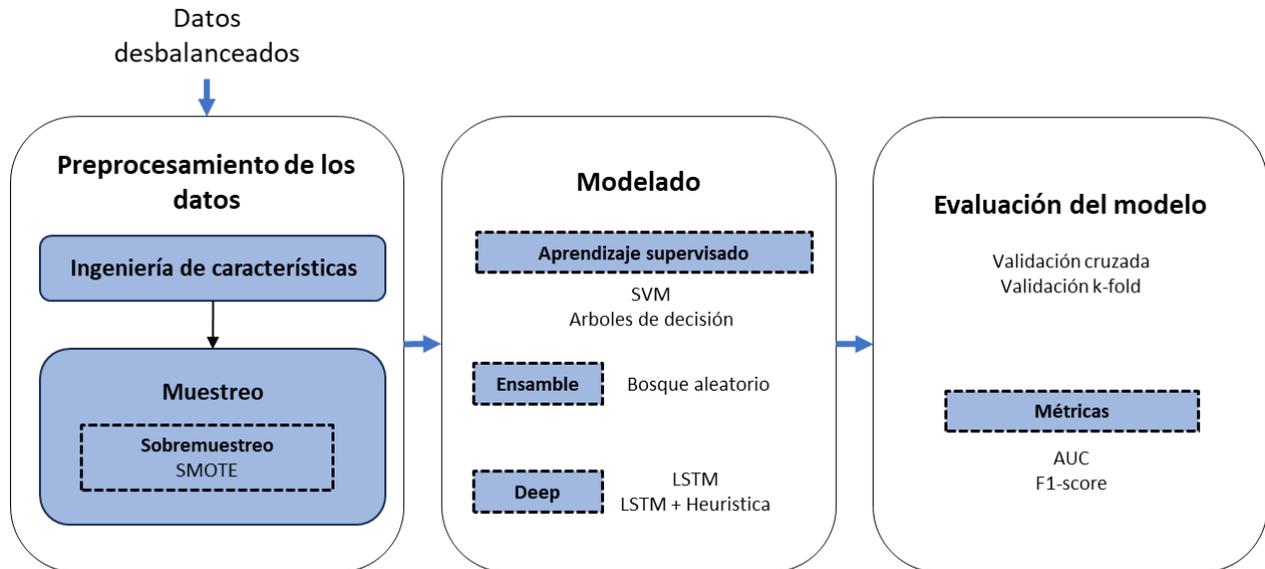


Ilustración 7: Flujo aprendizaje automático para la predicción y el análisis de abandono, (editado de [14], [16], [23])

3. Resultados y análisis

- 3.1. Información básica sobre los datos empleados
- 3.2. Análisis exploratorio de los datos
- 3.3. Aplicación de los modelos y análisis de los resultados

4. Conclusiones

xxxx

5. Recomendaciones

Xxxx

Referencias

- [1] D. Anil Kumar and V. Ravi, "Predicting credit card customer churn in banks using data mining," *International Journal of Data Analysis Techniques and Strategies*, vol. 1, no. 1, pp. 4-28, 2008, doi: 10.1504/IJDATS.2008.02002.
- [2] B. He, Y. Shi, Q. Wan, and X. Zhao, "Prediction of Customer Attrition of Commercial Banks based on SVM Model," *Procedia Comput Sci*, vol. 31, pp. 423-430, 2014, doi: 10.1016/j.procs.2014.05.286.
- [3] V. Agarwal, S. Taware, S. A. Yadav, D. Gangodkar, A. L. N. Rao, and V. K. Srivastav, "Customer - Churn Prediction Using Machine Learning," in *Proceedings of International Conference on Technological Advancements in Computational Sciences, ICTACS 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 893-899. doi: 10.1109/ICTACS56270.2022.9988187.
- [4] D. AL-Najjar, N. Al-Rousan, and H. AL-Najjar, "Machine Learning to Develop Credit Card Customer Churn Prediction," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 4, pp. 1529-1542, Dec. 2022, doi: 10.3390/jtaer17040077.
- [5] H. Risselada, P. C. Verhoef, and T. H. A. Bijmolt, "Staying Power of Churn Prediction Models," *Journal of Interactive Marketing*, vol. 24, no. 3, pp. 198-208, Aug. 2010, doi: 10.1016/j.intmar.2010.04.002.
- [6] M. Valenzuela *et al.*, *La reinención financiera en la era digital*, Asobancaria., vol. 1. Asobancaria, 2022.
- [7] L. E. Arango, L. Cardona-Sosa, and N. Pedraza-Jiménez, "The use of credit cards among low- and middle-income individuals in Colombia and the channels of monetary policy," *Econ Model*, vol. 94, pp. 150-169, Jan. 2021, doi: 10.1016/j.econmod.2020.09.018.
- [8] DANE, "Demografía y población." Accessed: Mar. 31, 2023. [Online]. Available: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>
- [9] Asobancaria, "Informe tarjetas de crédito 2022," Asobancaria. Accessed: Apr. 11, 2023. [Online]. Available: <https://www.asobancaria.com/informe-tarjetas-credito/>
- [10] Anusmita Bose and K. T. Thomas, "A Comparative Study of Machine Learning Techniques for Credit Card Customer Churn Prediction," in *Book cover Book cover Computer Networks and Inventive Communication Technologies*, vol. 141, Singapore: Springer, 2022, pp. 295-307.
- [11] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," *Financal Innovation*, vol. 2, no. 1, p. 10, Dec. 2016, doi: 10.1186/s40854-016-0029-6.
- [12] S. F., "Machine-Learning Techniques for Customer Retention: A Comparative Study," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, pp. 273-281, 2018, doi: 10.14569/IJACSA.2018.090238.
- [13] E. Ascarza *et al.*, "In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions," *SSRN Electronic Journal*, 2017, doi: 10.2139/ssrn.2903548.
- [14] L. Geiler, S. Affeldt, and M. Nadif, "A survey on machine learning methods for churn prediction A survey on machine learning methods for churn prediction A survey on machine learning methods for churn prediction," *Int.J Data Sci Anal*, vol. 2022, no. 3, pp. 217-242, doi: 10.1007/s41060-022-00312-5i.
- [15] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes," *Appl Soft Comput*, vol. 137, p. 110103, Apr. 2023, doi: 10.1016/j.asoc.2023.110103.
- [16] L. Geiler, S. Affeldt, and M. Nadif, "An effective strategy for churn prediction and customer profiling," *Data Knowl Eng*, vol. 142, Nov. 2022, doi: 10.1016/j.datak.2022.102100.

- [17] D. F. Benoit and D. Van den Poel, "Improving customer retention in financial services using kinship network information," *Expert Syst Appl*, vol. 39, no. 13, pp. 11435–11442, Oct. 2012, doi: 10.1016/j.eswa.2012.04.016.
- [18] X. Miao and H. Wang, "Customer Churn Prediction on Credit Card Services using Random Forest Method," 2022.
- [19] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by logistic regression and decision tree," *Expert Syst Appl*, vol. 38, no. 12, pp. 15273–15285, Nov. 2011, doi: 10.1016/j.eswa.2011.06.028.
- [20] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Appl Soft Comput*, vol. 24, pp. 994–1012, Nov. 2014, doi: 10.1016/j.asoc.2014.08.041.
- [21] G. XIA and W. JIN, "Model of Customer Churn Prediction on Support Vector Machine," *Systems Engineering - Theory & Practice*, vol. 28, no. 1, pp. 71–77, Jan. 2008, doi: 10.1016/S1874-8651(09)60003-X.
- [22] A. Sharma and P. Kumar Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services," *Int J Comput Appl*, vol. 27, no. 11, pp. 26–31, 2011, doi: 10.5120/3344-4605.
- [23] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Syst Appl*, vol. 39, no. 1, pp. 1414–1425, Jan. 2012, doi: 10.1016/j.eswa.2011.08.024.
- [24] X. Yu, S. Guo, J. Guo, and X. Huang, "An extended support vector machine forecasting framework for customer churn in e-commerce," *Expert Syst Appl*, vol. 38, no. 3, pp. 1425–1430, 2011, doi: 10.1016/j.eswa.2010.07.049.
- [25] C.-S. Lin, G.-H. Tzeng, and Y.-C. Chin, "Combined rough set theory and flow network graph to predict customer churn in credit card accounts," *Expert Syst Appl*, vol. 38, no. 1, pp. 8–15, Jan. 2011, doi: 10.1016/j.eswa.2010.05.039.
- [26] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert Syst Appl*, vol. 34, no. 1, pp. 313–327, Jan. 2008, doi: 10.1016/j.eswa.2006.09.038.
- [27] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Computing Surveys*, vol. 49, no. 2. Association for Computing Machinery, Aug. 01, 2016. doi: 10.1145/2907070.
- [28] T. M. Khoshgoftaar, C. Seiffert, J. Van Hulse, A. Napolitano, and A. Folleco, "Learning with limited minority class data," in *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, Florida: IEEE, Dec. 2007, pp. 348–353. doi: 10.1109/ICMLA.2007.76.
- [29] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," 2002.
- [30] V. Umayaparvathi and K. Iyakutti, "A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics," 2016. [Online]. Available: <http://www.fuqua.duke.edu/centers/crm/index.html>
- [31] O. Akbarzadeh *et al.*, "Heating-Cooling Monitoring and Power Consumption Forecasting Using LSTM for Energy-Efficient Smart Management of Buildings: A Computational Intelligence Solution for Smart Homes," 2024.
- [32] C. Alzaman, "Forecasting and optimization stock predictions: Varying asset profile, time window, and hyperparameter factors," *Systems and Soft Computing*, vol. 5, Dec. 2023, doi: 10.1016/j.sasc.2023.200052.
- [33] A. Sharma and P. Kumar Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services," 2011.

- [34] T. A. Rashid, M. K. Hassan, M. Mohammadi, and K. Fraser, "Improvement of Variant Adaptable LSTM Trained With Metaheuristic Algorithms for Healthcare Analysis," in *Advanced Classification Techniques for Healthcare Analysis*, IGI Global, 2019, pp. 111–131. doi: 10.4018/978-1-5225-7796-6.ch006.
- [35] A. De Caigny, K. W. De Bock, S. Lessmann, C. G. Mena, and K. Coussement, "Churn Prediction with Sequential Data and Deep Neural Networks. A Comparative Analysis Churn Prediction with Sequential Data and Deep Neural Networks A Comparative Analysis *," 2019. [Online]. Available: <https://www.researchgate.net/publication/336056150>
- [36] N. Jajam and N. P. Challa, "Dynamic Behavior-Based Churn Forecasts in the Insurance Sector," *Computers, Materials and Continua*, vol. 75, no. 1, pp. 977–997, 2023, doi: 10.32604/cmc.2023.036098.
- [37] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.
- [38] S. Punia, K. Nikolopoulos, S. P. Singh, J. K. Madaan, and K. Litsiou, "Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail," *Int J Prod Res*, vol. 58, no. 16, pp. 4964–4979, Aug. 2020, doi: 10.1080/00207543.2020.1735666.
- [39] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur J Oper Res*, vol. 270, no. 2, pp. 654–669, Oct. 2018, doi: 10.1016/j.ejor.2017.11.054.
- [40] A. Mastakouris, G. Andriosopoulou, D. Masouros, P. Benardos, G.-C. Vosniakos, and D. Soudris, "Human worker activity recognition in a production floor environment through deep learning," *J Manuf Syst*, vol. 71, pp. 115–130, Dec. 2023, doi: 10.1016/j.jmsy.2023.08.020.
- [41] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Computation MIT*, vol. 31, pp. 1235–1270, 2019, doi: 10.1162/NECO.
- [42] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, "Metaheuristic algorithms: A comprehensive review," in *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, Elsevier, 2018, pp. 185–231. doi: 10.1016/B978-0-12-813314-9.00010-4.
- [43] S. Stajkowski, D. Kumar, P. Samui, H. Bonakdari, and B. Gharabaghi, "Genetic-Algorithm-Optimized Sequential Model for Water Temperature Prediction," *Sustainability*, vol. 12, no. 13, p. 5374, Jul. 2020, doi: 10.3390/su12135374.
- [44] J. H. Jaramillo, J. Bhadury, and R. Batta, "On the use of genetic algorithms to solve location problems," *Comput Oper Res*, vol. 29, no. 6, pp. 761–779, May 2002, doi: 10.1016/S0305-0548(01)00021-1.
- [45] Z. K. Gurgi, A. I. Ismael, and R. A. Mejeed, "The Comparing Between Genetic Algorithm and Neural Network to Compute of Three-Basic Solar Cell Parameters with Wide Range of Measured Temperature," *Journal of Operation and Automation in Power Engineering*, vol. 12, no. 2, pp. 134–141, Apr. 2024, doi: 10.22098/joape.2023.10704.1774.
- [46] I. Granado, L. Hernando, Z. Uriondo, and J. A. Fernandes-Salvador, "A fishing route optimization decision support system: The case of the tuna purse seiner," *Eur J Oper Res*, Jan. 2023, doi: 10.1016/j.ejor.2023.07.009.
- [47] IBM, "Conceptos básicos de ayuda de CRISP-DM." Accessed: Aug. 05, 2021. [Online]. Available: <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=dm-crisp-help-overview>

Apéndices

Anexos