# Uso de modelos predictivos para la asignación de metas de tarjetas de crédito en los Centros de Atención de la empresa Tuya S.A.

Leonardo Navas Calderón

Informe de práctica para optar al título de Ingeniero Industrial

Asesor
Olga Cecilia Úsuga Manco, Doctora en ciencias - Estadística

Universidad de Antioquia
Facultad de Ingeniería
Ingeniería Industrial
Medellín
2023

| | |
|---|---|
| **Cita** | (Navas Calderón, 2023) |
| **Referencia**<br><br>**Estilo APA 7 (2020)** | Navas Calderón, L. (2023). *Uso de modelos predictivos para la asignación de metas de tarjetas de crédito en los Centros de Atención de la empresa Tuya S.A.* [pregrado presencial]. Universidad de Antioquia, Medellín. |



Créditos a escenario de prácticas, personas, proyectos que aportaron al desarrollo de la práctica (interna y externamente: empresa y área de la empresa, grupo de investigación, proyecto, organización)

## Dedicatoria

Dedicado a cada persona que de alguna u otra forma me impulsó a seguir sobre todo en los momentos más difíciles; profesores y compañeros que ahora se han convertido en amigos de la vida.

En memoria de mi padre Alberto, sin el cual nunca habría emprendido este camino tan maravilloso donde el ingenio y la persistencia formaron parte de su ser y ahora del mío.

A mi madre, Josefina que con su infinita sabiduría y amor supo darme guía todos estos años, a mis hermanos por su continuo apoyo e inspiración, sin la ayuda de todos esto no sería posible.


## Agradecimientos

Muchas gracias a cada uno de mis profesores por su infinita paciencia al enseñar, a mis compañeros de clase de quienes aprendí más de lo que me hubiera imaginado.
Total agradecimiento con la compañía Tuya SA por abrirme sus puertas, a mis jefes y compañeros de trabajo por su confianza depositada.

**Tabla de contenido**

## Lista de tablas

## Lista de figuras

## Siglas, acrónimos y abreviaturas

| | |
|---|---|
| **Catt** | Centro de Atención Tuya |
| **MLP** | Multi-layer Perceptron (Perceptrón Multicapa) |
| **SVR** | Support Vector Regressor |
| **RFR** | Random Forest Regressor |
| **MSE** | Error Cuadrático Medio |
| **R2** | Coeficiente de determinación |

**Resumen**

En la actualidad, la búsqueda constante de eficiencia y optimización en el sector servicios, especialmente en el financiero se ha convertido en la clave para estar siempre a la vanguardia en un sector que es cada vez más competitivo. Es en este contexto que se requiere la implementación de un modelo de metas de tarjetas de crédito para el área comercial de Tuya S.A como una estrategia clave para gestionar los servicios prestados por los Centros de Atención (Catts).

El enfoque se centra en primera medida, en la planificación precisa, para lo cual se llevó a cabo un análisis exploratorio de los datos brindados por el equipo de Desarrollo Comercial de la compañía de forma preliminar a la implementación de los modelos predictivos, una ejecución efectiva y el monitoreo de la calidad de los modelos predictivos, en este caso en particular usando los modelos de Support Vector Regressor, Random Forest para regresión y modelos utilizando redes neuronales. Finalmente, dicha implementación se traduce en una mejor utilización de los recursos que dispone la compañía ahorrando tiempo al automatizar un proceso que toma varios días en llevarse a cabo.

*Palabras clave*: asignación de metas, catt, tarjetas de crédito, modelos predictivos, random forest regressor, support vector regressor, mlp regressor.

**Abstract**

Nowadays, the constant search for efficiency and optimisation in the service sector, especially in the financial sector, has become the key to always being at the forefront in an increasingly competitive sector. It is in this context that the implementation of a credit card target model for the commercial area of Tuya S.A. is required as a key strategy to manage the services provided by the Customer Service Centres (Catts).

The approach focuses firstly on accurate planning, for which an exploratory analysis of the data provided by the company's Business Development team was carried out preliminary to the implementation of the predictive models, effective execution and monitoring of the quality of the predictive models, in this particular case using Support Vector Regressor models, Random Forest for regression and models using neural networks. Ultimately, such implementation translates into better utilisation of the company's available resources, saving time by automating a process that takes several days to complete.

*Keywords*: goal assignment, catt, credit cards, predictive models, random forest regressor, support vector regressor, mlp regressor.

## Introducción

En el mundo cada vez más competitivo de las empresas financieras, la toma de decisiones estratégicas es esencial para el éxito de la organización, uno de los desafíos cruciales que enfrenta una empresa de este sector es la optimización de sus servicios y productos, como la oferta de tarjetas de crédito, para satisfacer las necesidades de sus clientes y garantizar la rentabilidad del negocio.

Este trabajo busca presentar una propuesta de implementación de varios modelos de aprendizaje supervisado y de series de tiempo para las metas de colocación de tarjetas de crédito para los Centros de Atención (Catts) de Tuya SA. La correcta distribución de las metas para cada ejecutivo y Catt por mes es esencial para garantizar la satisfacción de los clientes, al mismo tiempo que se optimizan los procesos internos y se promueve la eficiencia de los ejecutivos comerciales y los colaboradores.

Esta implementación se llevará a cabo mediante el análisis de series de tiempo, un modelo de bosques aleatorios, y un vector de soporte para regresión usando el lenguaje de programación Python con el fin de observar diferentes escenarios de pronóstico y elegir el modelo adecuado según la naturaleza de los datos de cada Catt usando la metodología Crisp DM.

# 1 Objetivos

## 1.1 Objetivo general

Proponer la implementación de modelos predictivos para la asignación de metas de tarjetas de crédito para los ejecutivos comerciales en los Centros de Atención administrados por Tuya S.A a nivel nacional.

## 1.2 Objetivos específicos

1. Identificar datos históricos para la predicción de la oferta de tarjetas de crédito en cada Centro de Atención - Catt.
2. Ajustar modelos predictivos teniendo en cuenta las variables influyentes para lograr una mayor precisión en las predicciones.
3. Evaluar la calidad de los modelos predictivos propuestos haciendo uso de métricas de desempeño.
4. Implementar un sistema de toma de decisiones basado en los modelos predictivos que recomiende la cantidad óptima de tarjetas de crédito a ofrecer en cada Centro de Atención - Catt.

# 2 Marco teórico

Dentro de los modelos predictivos, los modelos de series de tiempo y de aprendizaje supervisados han sido ampliamente utilizados desde hace algunos años, ya que han demostrado ser de gran utilidad para la toma de decisiones debido a una implementación sencilla y de fácil comprensión. Además, "disponen de unas medidas que permiten valorar la calidad de las estimaciones, entre las que se encuentran el coeficiente de correlación, el coeficiente de determinación y el error estándar de estimación, entre otros" (Sarmiento, 2008, p. 42).

En este caso, se utilizaron tres modelos: Random Forest, Support Vector Regressor, y un modelo basado en redes neuronales usando un Perceptrón Multicapa (MLP Regressor).

## 2.1 Support Vector Regressor (SVR):

El Suport Vector Regressor (SVR) es una variante de una máquina de soporte vectorial (SVM) ampliamente utilizada para dar solución a problemas de regresión cuyo objetivo central es encontrar una función que sea capaz de aproximar y predecir de manera precisa los valores de salida a partir de un conjunto de datos de entrada minimizando al mismo tiempo la cantidad de errores de predicción, en este caso el de la colocación de tarjetas de crédito en los Catts a nivel nacional.

Se da un conjunto de datos de entrenamiento $\{x_n,\ y_n\}_{n=1}^{N},\ x_n \in R^d,\ y_n \in R$, donde $x_i$ es la entrada y $y_i$ es la salida, cuando el SVR más simple no se ajusta a los datos lo suficientemente bien, un SVR se representa como sigue en la Ecuación 1:

$$min\frac{1}{2}\left|\left|w\right|\right| + C\sum_{i=1}^{l}(\xi_l\ + \xi_l^*) \qquad (1)$$

Sujeto a

$$y_j - f(x_i,\ \omega) \leq \varepsilon + \xi_i\ \ldots\ \ldots,\ l;$$
$$f(x_i,\ \omega) - y_i \leq \varepsilon + \xi_l^*, \cdots, \ldots,\ l;$$
$$\xi_i \geq 0,\ \xi_i^* \geq 0,\ i = 1,2, \ldots,\ l;$$

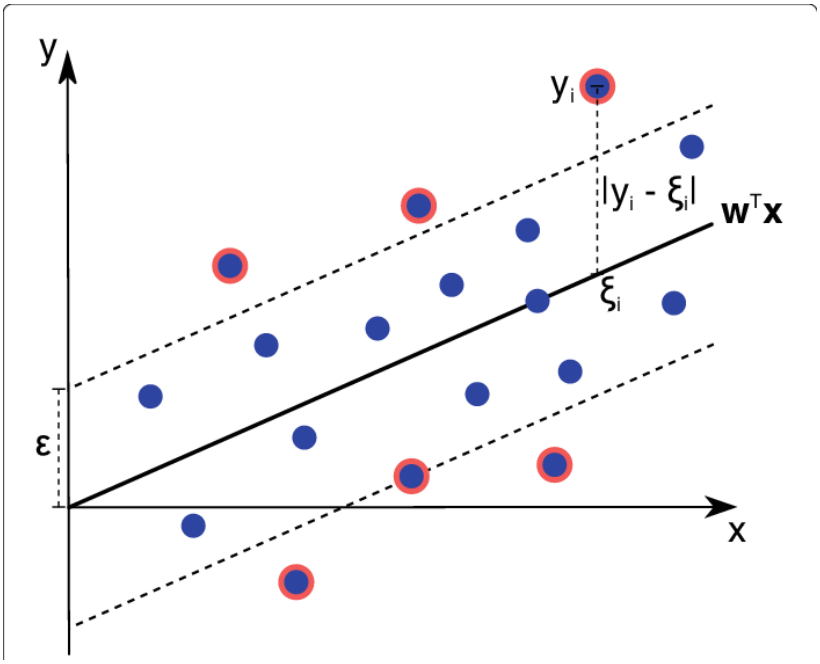
Donde ‖ω‖ representa el factor relacionado con la complejidad del modelo; C es el grado del castigo para las muestras que superen el error; ε es la función de pérdida insensible cuyo valor

afecta al número de vectores de soporte; $\xi_i$ y $\xi_i^*$ son variables de holgura que indican en qué medida la muestra se desvía de la zona insensible (Tan et al., 2019).

A continuación, en la Figura 1 se puede observar la estructura del SVR representada por la función $W^TX$. Se muestra el "tubo insensible a errores" (ε -insensitive tube) en color gris. Este tubo es una región alrededor de la línea de regresión donde las desviaciones de las predicciones respecto a los valores reales, dentro de un cierto límite (ε), no contribuyen a la función de pérdida. La función de pérdida penaliza las desviaciones fuera de este tubo, pero no las penaliza si están dentro de él, $\xi_i = w^Tx_i$ representa el valor objetivo predicho de $x_i$; $y_i$ hace referencia al valor actual del objetivo (Rosenbaum et al., 2013).

**Figura 1**

*Estructura de un SVR*




Nota. Fuente https://www.researchgate.net/figure/Support-vector-regression-SVR-Illustration-of-an-SVR-regression-function-represented_fig12_248396465

**Hiperparámetros:**

El modelo SVR utiliza los siguientes hiperparámetros descritos a continuación; el hiperparámetro 'C' de regulación controla la penalización de los errores en función de la pérdida, un valor alto de C significa un ajuste más preciso del conjunto de entrenamiento. Por otro lado, el

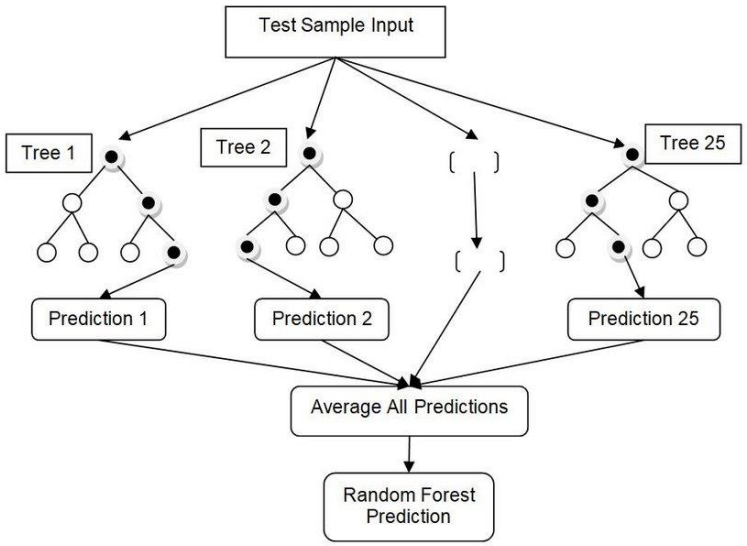
'Kernel' o núcleo define la función a utilizar, por ejemplo, lineal, polinomico, radial o sigmoidal. Entre tanto, el 'Degree' se refiere al grado del polinomio a utilizar en la función del kernel. Por último, 'Epsilon' se refiere al margen de error aceptable para la regresión.

**2.2 Bosque aleatorio (Random Forest):**

Un bosque aleatorio o random forest regressor es una variación de los árboles de decisión, este modelo "funciona construyendo una multitud de árboles de decisión en el momento del entrenamiento y generando una predicción media de los árboles individuales" (Pangarkar et al., 2020), cada árbol extrae una muestra de manera aleatoria del conjunto de datos original y genera divisiones que evitan el sobreajuste del modelo. En la Figura 2 se muestra la estructura de un Random Forest Regressor.

**Figura 2**

*Estructura del Random Forest Regressor*




Nota. Fuente https://www.researchgate.net/figure/Random-forest-structure-15_fig3_346411178

**Hiperparámetros:**

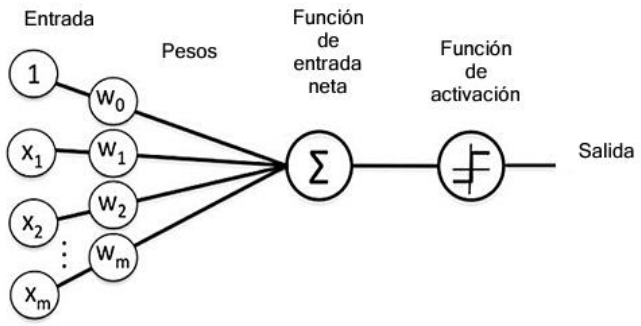
Los hiperparametos utilizados para el modelo Random Forest regressor se definen a continuación; el parámetro 'max_depth' hace referencia a la profundidad máxima de los árboles; el 'Min_samples_split' es el número mínimo de muestras requeridas para dividir un nodo interno,

por último, el ‘n_estimators’ hace referencia al número de árboles en el bosque (Contreras et al., 2021).

**2.3 Perceptrón Multicapa para regresión (MLP Regressor):**

Un MLP Regressor es una red neuronal artificial (RNA) la cual se define como un modelo computacional inspirado en la estructura y funcionamiento del cerebro humano, cada neurona artificial, se conecta a otra, tiene un peso asociado y un umbral como se aprecia en la Figura 3. Las MLP son las más antiguas y usadas redes neuronales, están formadas por una capa de entrada (Input layer), una capa o varias capas ocultas (Hidden layer) y una capa de salida (Output layer), este modelo optimiza el error cuadrático medio (MSE) usando el gradiente estocástico descendente (IBM, 2021).

**Figura 3**
*Red neuronal*

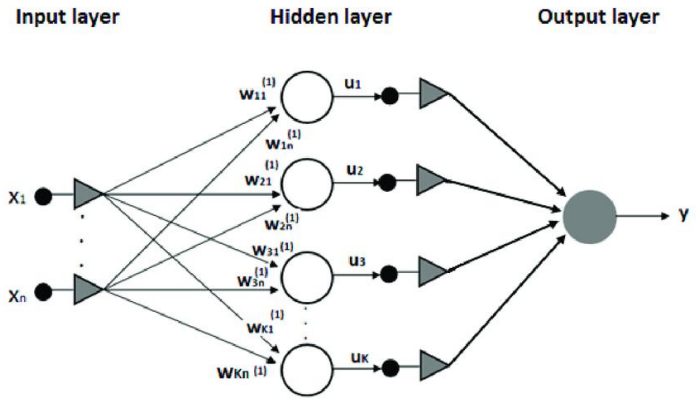



Nota. Fuente https://wiki.pathmind.com/neural-network

El MLP Regressor se utiliza para resolver problemas de regresión, donde el objetivo es predecir un valor numérico continuo en función de una serie de variables de entrada, en la Figura 4 se puede observar la estructura de un Perceptrón Multicapa (MLP).

**Figura 4**

*MLP Regressor*




Nota. Fuente https://www.researchgate.net/figure/Example-of-multilayer-perceptron-MLP-network-for-regression-analysis-here-n-13-and_fig1_347244673

**Hiperparámetros:**

Se definen los hiperparámetros del MLPR a continuación:

El parámetro 'activation' hace referencia a la función de activación utilizada en las capas ocultas de la red neuronal, por ejemplo, relu (Rectified Linear Unit), tanh (tangente hiperbólica) y sigmoid (función sigmoide) (Bae et al., 2019).

Para medir la calidad de un estimador o predictor de un modelo utilizado se tienen en cuenta las métricas de desempeño como el coeficiente de determinación (R2), el error cuadrático medio (MSE) y el error cuadrático medio de la raíz (RMSE), a continuación, se hace una descripción de cada uno de las métricas utilizadas:

Se definen las métricas usadas para evaluar el desempeño a continuación en la (Ecuación 2, Ecuación 3 y Ecuación 4) según (Sarmiento, 2008):

**Error cuadrático medio o varianza residual (MSE):** es una medida de error promedio de las estimaciones. La variable $\bar{y}$ es la media de las estimaciones. Se define como:

$$S^2_{r\left(\frac{y}{x}\right)} = \frac{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}{N} \qquad (2)$$

**Coeficiente de determinación $R^2$:** se define como la proporción de la varianza explicada respecto de la varianza total de la variable explicada por la regresión. Refleja la bondad del ajuste de un modelo a la variable *Y* que pretender explicar:

$$S_y^2 = \frac{S_{yhat}^2 + S_{r\left(\frac{y}{x}\right)}^2}{N} \qquad (3)$$

**Error cuadrático medio de la raíz (RMSE):**

El RMSE indica la cantidad de error que hay entre los valores predichos por el modelo y los valores observados, el RMSE de un modelo predictivo se define como:

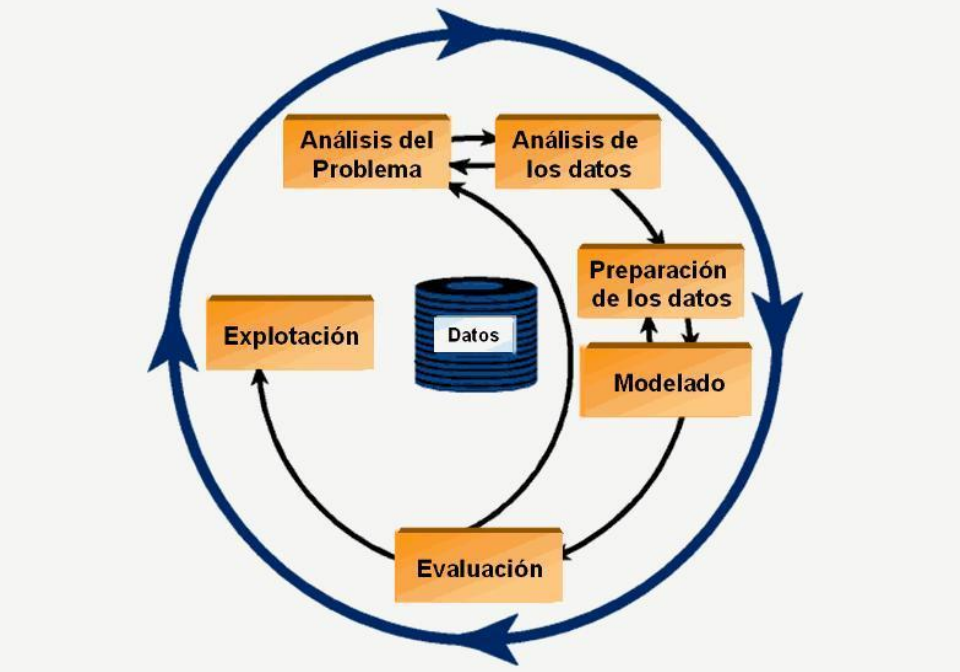
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{obs,i} - X_{model,i}\right)^2}{n}} \quad (4)$$

Donde $X_{obs}$ son los valores observados y $X_{model}$ son los valores modelados en el momento i.

## 3 Metodología

El presente trabajo se desarrolló siguiendo la metodología CRISP-DM (CRoss-Industry Standard Process for Data Mining), para una mayor comprensión de esta metodología, se hará un despliegue paso a paso para lo cual se tiene la siguiente estructura que se muestra en la Figura 5.

**Figura 5**
*Estructura metodología CRISP-DM*




Nota. Fuente https://www.researchgate.net/figure/Fases-del-proceso-KDD-segun-la-metodologia-CRISP-DM-Primeramente-se-debe-estudiar-el_fig1_233426470

Esta metodología se compone de seis fases con flechas que indican el flujo y dependencias entre ellas, cabe recalcar que las secuencias no son estrictas. Las fases comprenden el análisis del problema y el entendimiento del negocio, la comprensión de los datos, la preparación de los datos, el modelado, la evaluación y, por último, la explotación (Ordoñes et al., 2011), es especialmente útil para planificar y explicar la ejecución de un proyecto, las fases para este proyecto se describen a continuación:

**Fase I: Análisis del problema y el entendimiento del negocio**

En esta primera etapa se realizó una exploración inicial del problema a resolver, en este caso se trató de la asignación de metas de tarjetas de crédito a ofrecer para los ejecutivos comerciales en los Centros de Atención a nivel nacional de la empresa Tuya S.A. Se realizó un levantamiento de las bases de datos mediante el uso de la herramienta de relacionamiento de bases de datos Squirrel SQL. Posteriormente, se describieron las variables relacionadas a la solución del problema en la Tabla 1 con base en la opinión de un experto del área de Desarrollo Comercial de la compañía.

**Tabla 1**
*Variables de los modelos de aprendizaje automático*

| Variable | Definición | Tipo |
|---|---|---|
| Indtt | Indicador de tarjetas de crédito por Catt por mes. | Independiente |
| r_clixh | Número de clientes que se aprueban por hora. | Explicativa |
| porcaprobacion | Tasa de capturas*. | Explicativa |
| diashabiles | Cantidad de días hábiles en el mes. | Explicativa |
| MesPromo | Variable binaria: 1 si hay promociones ese mes; 0 de lo contrario. | Explicativa |
| ind_capturas | Indicador de capturas de tarjetas de crédito: # de capturas por hora. | Explicativa |

** Una captura es una consulta de la cedula de ciudadanía de un potencial cliente en el motor de originación (PCO); lo cuales son motores de calificación crediticia.*

**Fase II: Análisis de los datos**

En esta segunda etapa se incluyó un análisis de los datos para comprender su naturaleza, calidad y estructura para lo cual se realiza una exploración inicial de los datos históricos en Python, los datos están estructurados con una temporalidad mensual a día de corte para 32 meses y para los 112 Catts a nivel nacional, los datos inician el día 31 de enero del 2021 y finalizan el día 30 de agosto del año 2023.

**Fase III: Preparación de los datos**

En esta fase se realizó una inspección de los datos recabados, detectando datos duplicados, nulos o faltantes y determinando con base en las recomendaciones de un experto en el negocio de la empresa Tuya S.A qué tratamiento se les darían a dichos datos, para lo cual se realizó un promedio de los últimos tres meses de los datos concernientes a dicho valor de la variable nula o faltante.

**Fase IV: Modelado**

Es en esta fase donde se seleccionan y aplican distintas técnicas de modelado con el fin de construir y evaluar los modelos de Suport Vector Regressor, Random Forest y un modelo utilizando Redes neuronales que puedan resolver el problema planteado para la estimación de metas para los ejecutivos comerciales en Tuya S.A.

**Fase V: Evaluación**

Usando las métricas de desempeño se evalúan los modelos construidos, a través del uso del R2 para el cual un valor alto significa un mejor ajuste del modelo, el MSE para el cual un valor bajo significa una menor desviación de los datos, al igual que el RMSE el cual es la medida de la dispersión de los errores de predicción, y cuanto menor sea mejor será el ajuste del modelo, se realizan comparaciones y se selecciona el mejor modelo dependiendo de la naturaleza de los datos de los Catts.

**Fase VI: Implementación**

En esta fase se pretende implementar el o los modelos seleccionados según las necesidades de cada Catt, se realiza un monitoreo y alimentación con datos nuevos cada mes, estos son proporcionados por el equipo de Desarrollo Comercial de Tuya S.A., sin embargo, para el alcance de este proyecto no se realizará la implementación de los modelos debido a que al momento de finalizar este trabajo, se espera llevar a cabo una revisión exhaustiva y posterior colaboración con el equipo de Analítica de Datos de la empresa para llevar a cabo un piloto del modelo de metas de tarjetas de crédito en los Catts en las regiones de Bogotá y poblaciones cercanas.

# 4 Resultados y análisis

## 4.1 Exploración de los datos

A continuación, se describen los datos obtenidos con los cuales se va a desarrollar el presente trabajo donde se tienen datos históricos para el análisis con temporalidad mensual a día de corte para 32 meses y para los 112 Catts a nivel nacional, los datos inician el día 31 de enero del 2021 y hasta el día 30 de agosto del año 2023, a continuación, en la Figura 6 se observa la estructura de los datos en las primeras cinco filas, en la Figura 7 se evidencia el resumen estadístico de los datos para cada una de las variables numéricas y en la Figura 8 se observa la base de datos depurada donde no se encuentran datos nulos o faltantes:

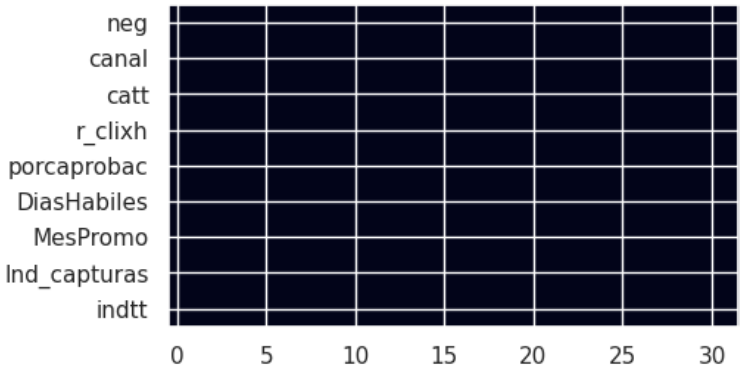
**Figura *6***
*Visualización de los datos para el Catt Almacén Alameda*

| | neg | canal | catt | fecha_corte | r_clixh | porcaprobac | DiasHabiles | MesPromo | Ind_capturas | indtt |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | E | 305057 | ALMACÉN ALAMEDA | 2021-01-31 | 6.27 | 0.1015 | 24 | 0 | 1.66 | 39.25 |
| **1** | E | 305057 | ALMACÉN ALAMEDA | 2021-02-28 | 5.74 | 0.1281 | 24 | 1 | 2.52 | 66.74 |
| **2** | E | 305057 | ALMACÉN ALAMEDA | 2021-03-31 | 5.33 | 0.1100 | 26 | 1 | 2.55 | 66.16 |
| **3** | E | 305057 | ALMACÉN ALAMEDA | 2021-04-30 | 5.94 | 0.1141 | 24 | 0 | 1.60 | 43.65 |
| **4** | E | 305057 | ALMACÉN ALAMEDA | 2021-05-31 | 5.77 | 0.1390 | 24 | 0 | 1.82 | 52.19 |

**Figura *7***
*Resumen estadístico de los datos*

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **canal** | 32.0 | 305057.000000 | 0.000000 | 305057.000 | 305057.0000 | 305057.00000 | 305057.000000 | 305057.0000 |
| **r_clixh** | 32.0 | 8.423438 | 1.743357 | 5.330 | 7.0500 | 8.68500 | 9.845000 | 11.0700 |
| **porcaprobac** | 32.0 | 0.165091 | 0.043761 | 0.097 | 0.1387 | 0.16175 | 0.179275 | 0.2889 |
| **DiasHabiles** | 32.0 | 24.656250 | 0.865443 | 23.000 | 24.0000 | 24.00000 | 25.000000 | 26.0000 |
| **MesPromo** | 32.0 | 0.500000 | 0.508001 | 0.000 | 0.0000 | 0.50000 | 1.000000 | 1.0000 |
| **Ind_capturas** | 32.0 | 1.917813 | 0.336277 | 1.230 | 1.6575 | 1.89500 | 2.087500 | 2.5500 |
| **indtt** | 32.0 | 63.311250 | 22.177138 | 25.170 | 45.7925 | 60.40000 | 70.352500 | 125.9200 |

**Figura *8***
*Visualización de la base de datos depurada*




En la exploración de los datos se encontraron como principales hallazgos que las bases de datos tienen una buena calidad, ya que no tienen mayores problemas de datos faltantes o nulos y estos fueron resueltos de manera satisfactoria.

### 4.2 Modelado

Se crean varios modelos predictivos; un modelo de Suport Vector Regression (SVR), un Random Forest Regressor, además, un modelo usando redes neuronales con múltiples capas (MLPR) con nueve pasos hacia atrás (look back) utilizando las variables de la Tabla 1 para cada uno de los Catts, adicionalmente, se les hace un ajuste de hiperparámetros con el método de búsqueda en rejilla (grid search) con el fin de encontrar el mejor modelo posible con los datos disponibles, se resumen a continuación en la Tabla 2 los mejores hiperparámetos encontrados:

**Tabla 2**
*Mejores Hiperparámetros de los modelos*

| Modelo | Hiperpárametros |
|---|---|
| SVR | ‘C’= 1; ‘Degree’ = 2; ‘Epsilon’ = 0.1; ‘Kernel’ = lineal |
| RFR | ‘Max_depth’ = 5; ‘Min_samples_split’ = 2; ‘n_estimators’ = 100 |
| MLP Regressor | ‘Activation’ = tanh*; ‘solver’ = adam |

*tanh: tangente hiperbólica

En la fase de modelado se encontró que los modelos SVR y RFR resultaron con unos mayores ajustes y predicciones en general, ya que presentan una mayor adaptación a las variaciones de los datos de cada Catt.

### 4.3 Evaluación de los modelos

Como resultado de la construcción, modelado, entrenamiento y testeo de los modelos, se obtiene el siguiente resumen de las métricas de desempeño (R2, MSE y RMSE) de los modelos utilizados como se muestra en la Tabla 3, además, se observarán las predicciones de manera gráfica como se muestra en las siguientes figuras (Figura 25, Figura 26 y Figura 27) con el fin de realizar una comparación frente los datos reales observados.

Por un lado, la interpretación de las métricas da como resultado lo siguiente:

**R2**

El coeficiente de determinación R2 es una medida estadística que indica la proporción de la variabilidad de la variable dependiente que se explica por una o varias variables independientes. Su valor varía de cero a uno, donde cero indica que el modelo no explica la variabilidad y uno que el modelo explica toda la variabilidad. En este caso, el modelo que mejor explica la variabilidad es el SVR con un 0.78, lo que indica que el indicador de tarjetas de crédito (Indtt) puede explicarse aproximadamente en un 78%.

**MSE**

El MSE (Error Cuadrático Medio) es una medida de qué tan bien un modelo de regresión se ajusta a los datos reales, un MSE más bajo indica un mejor ajuste del modelo a los datos, mientras que un MSE más alto indica un peor ajuste o mayores errores en las predicciones.

En el caso del modelo SVR, como se puede observar en la Tabla 3, el R2 es el mayor lo cual favorece al modelo por encima de los demás, y el modelo MLPR resulta con el peor desempeño en esta métrica. Aunque el modelo MLP Regressor posee un MSE y RMSE bajos en comparación con los otros dos modelos, el R2 indica que existe un ajuste muy pobre de los datos para este modelo, por lo tanto, como resultado el modelo predictivo SVR se escoge por encima de los demás modelos en el centro de servicio de estudiado.

**Tabla 3**
*Resumen de las métricas de los modelos*

| Modelo | R2 | MSE | RMSE |
|---|---|---|---|
| SVR | 0.78 | 79.34 | 8.91 |
| RFR | 0.62 | 143.63 | 11.98 |
| MLP Regressor | 0.37 | 0.01 | 0.1 |

Por último, se realizan comparaciones de los valores predichos contra los valores observados de manera gráfica para observar el comportamiento de los modelos a nivel general, donde se pueden traducir los valores de la Tabla 3 en un gráfico de líneas como el que se observa en la Figura 9, Figura 10 y Figura 11, donde se tienen como principales resultados, por un lado, que el modelo MLP Regressor es quien presenta los valores más bajos de MSE, lo que indica que las predicciones pueden ser cercanas a los valores reales, sin embargo, un R2 bajo en comparación con los demás modelos indica que el modelo no esta brindando una buena explicación de la variabilidad de los datos. Por otro lado, el modelo SVR es el que brinda una mejor explicación de la variabilidad de los datos al tener el mayor R2 de los tres modelos estudiados.

**Figura 9**
*Predicciones del modelo Support Vector Regressor*

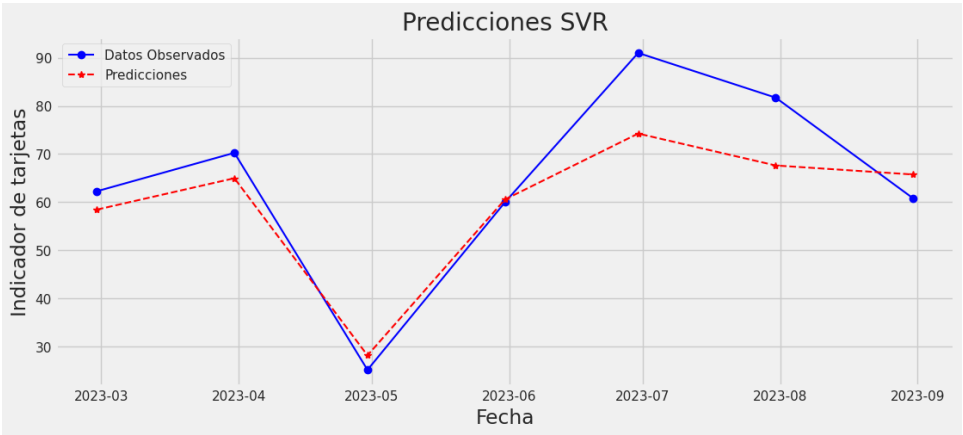



**Figura 10**
*Predicciones del modelo Random Forest*

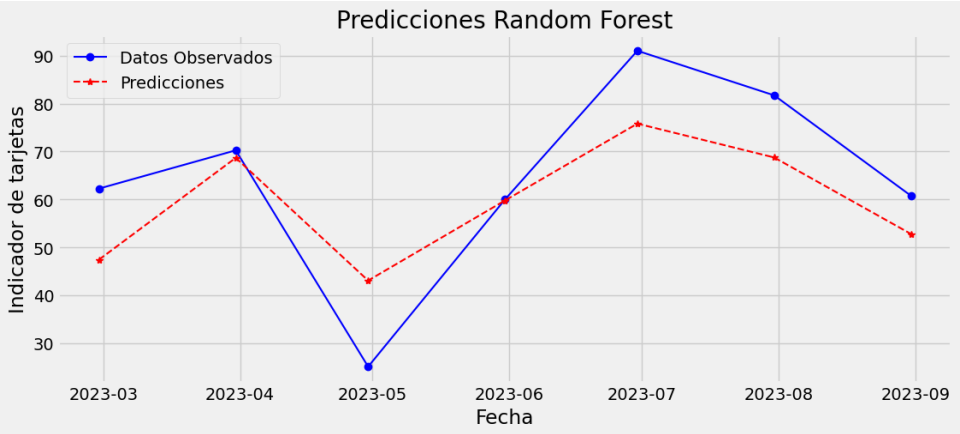
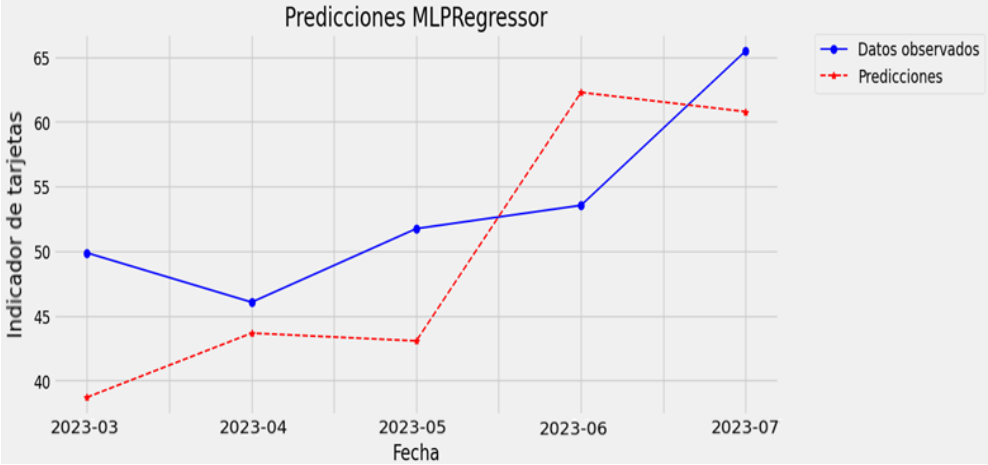



**Figura 11**
*Predicciones del modelo MLPR*



# 5 Conclusiones

Al finalizar este trabajo se pueden llevar a cabo las siguientes conclusiones mediante el uso de los modelos predictivos:

En las fases de análisis del problema y de los datos se realizó una exploración de los datos que se tenían disponibles en la compañía con miras a resolver el problema de la asignación de metas de las tarjetas de crédito para los ejecutivos comerciales, ya que como escenario inicial se tiene que actualmente dicha asignación de metas se realiza de manera empírica. Por otro lado, se encontró que, aunque los datos disponibles no son abundantes, sirven como un primer acercamiento e insumo para posteriores implementaciones de modelos predictivos cada vez más acertados al alimentar las bases de datos ya construidas.

Por otro lado, en la fase de la preparación de los datos se realizó la limpieza de datos faltantes o nulos, que dicho sea de paso no representaron ningún inconveniente ya que se poseía inicialmente una base de datos bastante limpia, por lo tanto, esta fase se llevó a cabo sin mayores inconvenientes, ya que la empresa posee buenas prácticas en el manejo de la información de las bases de datos.

En la fase de modelado se crearon varios modelos que se pueden ajustar a las necesidades de cada Catt facilitando una posterior implementación, lo cual representa un gran avance para este proceso en la compañía, ya que actualmente se realiza de manera empírica empleando varios días a esta tarea cada mes.

Para la fase de evaluación de los modelos se tuvieron en cuenta diferentes métricas de desempeño como el coeficiente de determinación R2, MSE y RMSE las cuales son importantes para la comparación entre los distintos modelos en términos de precisión y detección de problemas, entre otros aspectos relevantes.

Por último, estos modelos estarán prestos a implementarse en un futuro de acuerdo a las decisiones tomadas por el equipo de Desarrollo Comercial de la compañía, lo cual le brindará en términos generales un ahorro de tiempo y recursos humanos muy importantes.

## Referencias

Abollado, J. U., Mira, M., & Titulaci, M. (n.d.). *Aplicación de técnicas de simulación al estudio de modelos multicapa*.

Assessment of the Different Machine Learning Models for Prediction of Cluster Bean (Cyamopsis tetragonoloba L. Taub.) Yield - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Sample-random-forest-regression-tree_fig1_343992982 [accessed 15 Jan, 2024]

Bae, J., Han, J., Lee, D., Yang, J., Kim, J., Lim, K., Neff, J., & Jang, W. (2019). Evaluation of Sediment Trapping Efficiency of Vegetative Filter Strips Using Machine Learning Models. *Sustainability*, *11*, 7212. https://doi.org/10.3390/su11247212

Contreras, P., Orellana-Alvear, J., Muñoz, P., Bendix, J., & Célleri, R. (2021). Influence of Random Forest Hyperparameterization on Short-Term Runoff Forecasting in an Andean Mountain Catchment. *Atmosphere*, *12*, 238. https://doi.org/10.3390/atmos12020238

IBM. (2021). *¿Qué son las redes neuronales?* https://www.ibm.com/es-es/topics/neural-networks

Nicholson, C. (2019). *A Beginner's Guide to Neural Networks and Deep Learning.* https://wiki.pathmind.com/neural-network

Ordoñes, Y., Horacio, D., & Grass Boada, D. (2011). *Herramienta de Minería de Uso de la Web Aplicado a los Registros del Proxy*.

Pangarkar, D., Sharma, R., Sharma, A., & Sharma, M. (2020). Assessment of the Different Machine Learning Models for Prediction of Cluster Bean (Cyamopsis tetragonoloba L. Taub.) Yield. *Advances in Research*, 98–105. https://doi.org/10.9734/air/2020/v21i930238

Rosenbaum, L., Dörr, A., Bauer, M., Boeckler, F., & Zell, A. (2013). Inferring multi-target QSAR models with taxonomy-based multi-task learning. *Journal of Cheminformatics*, *5*, 33. https://doi.org/10.1186/1758-2946-5-33

Rudd, J., & Ray, H. (2020). An Empirical Study of Downstream Analysis Effects of Model Pre-Processing Choices. *Open Journal of Statistics*, *10*, 735–809. https://doi.org/10.4236/ojs.2020.105046

Sarmiento, E. M. (2008). *Predicción con series de tiempo y regresión.* 2, 36–58.

Tan, S., Ma, W., Hei, X., Xie, G., Chen, X., & Zhang, J. (2019). High Speed Train Axle Temperature Prediction Based on Support Vector Regression. *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2223–2227. https://doi.org/10.1109/ICIEA.2019.8833696