

Churn Modelling con Redes Neuronales

Roberto Torres

Proyecto II
Facultad de Ciencias
UNAM

rtorres.bello@ciencias.unam.mx

Enero 18, 2022

Índice

- 1 Customers Churn Prediction in Financial Institution Using Artificial Neural Network
 - Introducción
 - Ciencia de Datos
 - Intuición Algoritmo
 - Algoritmo

1 Customers Churn Prediction in Financial Institution Using Artificial Neural Network

- Introducción
- Ciencia de Datos
- Intuición Algoritmo
- Algoritmo

Introducción

- En este proyecto, se desarrolla un modelo predictivo que utiliza un perceptrón multicapa de la arquitectura de red neuronal artificial para predecir la rotación de clientes en una institución financiera.

Idea principal

A las empresas no les gusta perder clientes valiosos

Idea principal

A las empresas no les gusta perder clientes valiosos **en particular** a los bancos.



La pérdida de clientes es un problema fundamental para las empresas.

La pérdida de clientes es un problema fundamental para las empresas.

Se define como la pérdida de clientes porque se trasladan a la competencia.

Los clientes están directamente relacionados con las ganancias, las instituciones financieras deben evitar la pérdida de clientes mientras adquieren nuevos clientes.



Harvard Business Review cree que al reducir la tasa de deserción de clientes en un 5 %, las empresas pueden aumentar las ganancias entre un 25 % y un 85 %, mientras que Business Week cree que las ganancias aumentarán en un 140 %.



¿Qué es Churn Modelling?

Es un modelo predictivo que estima, a nivel de clientes individuales, la propensión (o susceptibilidad) que tienen a irse.



Es mucho más caro iniciar sesión en un nuevo cliente que mantener uno existente.



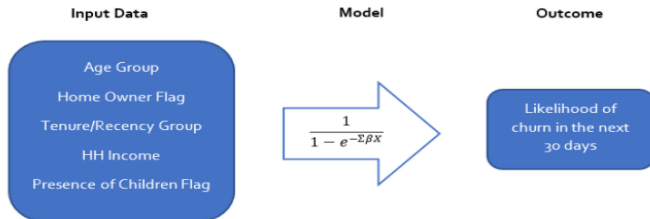
Es mucho más caro iniciar sesión en un nuevo cliente que mantener uno existente.



- La prevención de abandonos permite a las empresas desarrollar programas de fidelización y campañas de retención para mantener tantos clientes como sea posible.

¿Cómo entra la Ciencia de Datos

Las técnicas avanzadas de aprendizaje automático (ML) y ciencia de datos (DS) pueden aprender del comportamiento pasado del cliente y los factores desencadenantes externos que llevaron a la deserción y utilizar este aprendizaje para predecir la ocurrencia futura de un evento similar a una deserción.



Intuición

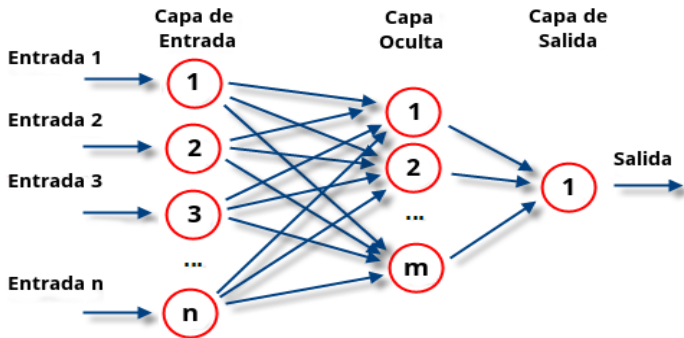
- Técnicamente, es un clasificador binario que divide a los clientes en dos grupos (clases): los que se van y los que no.



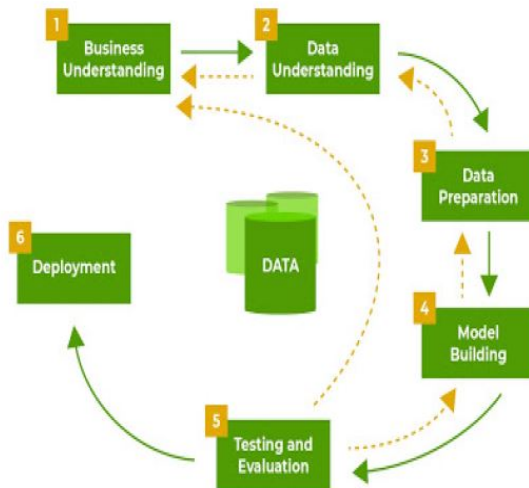
Es ventajoso para los bancos saber qué lleva a un cliente a la decisión de dejar la empresa.

Perceptrón Multicapa

Se hace uso del Perceptrón multicapa para predecir la rotación de clientes.

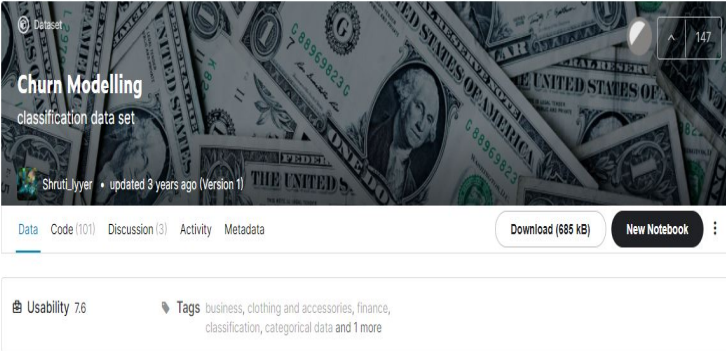


Flujo



Dataset: Bank Customer Churn Prediction

Se hará uso del dataset llamado Bank Customer Churn Prediction, este archivo se refiere a clientes de un banco que abandonaron la relación con el mismo.



The screenshot shows the Kaggle dataset page for "Churn Modelling" by Shruti_llyer. The background image is a collage of US dollar bills. The dataset title "Churn Modelling" is prominently displayed, followed by the subtitle "classification data set". The creator's name "Shruti_llyer" and the update date "updated 3 years ago (Version 1)" are shown. Below the title, there are tabs for "Data", "Code (101)", "Discussion (3)", "Activity", and "Metadata". The "Data" tab is selected. To the right of the tabs are buttons for "Download (685 kB)" and "New Notebook". At the bottom, the "Usability" score is 7.6, and the "Tags" include "business, clothing and accessories, finance, classification, categorical data and 1 more".

Dataset

Churn Modelling

classification data set

Shruti_llyer • updated 3 years ago (Version 1)

Data Code (101) Discussion (3) Activity Metadata

Download (685 kB) New Notebook

Usability 7.6

Tags business, clothing and accessories, finance, classification, categorical data and 1 more

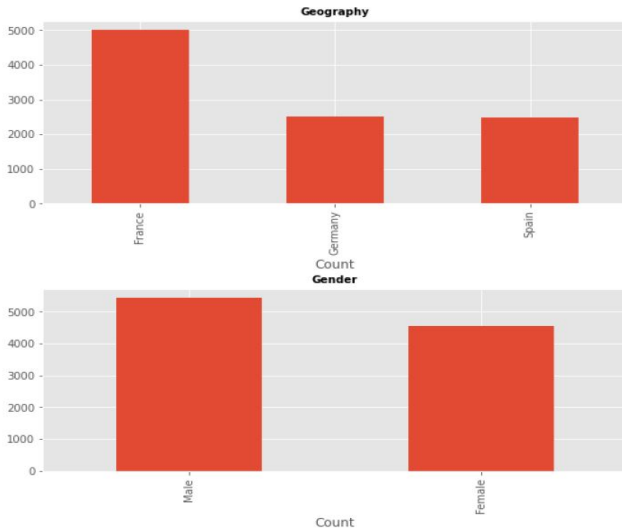
Algunos Features

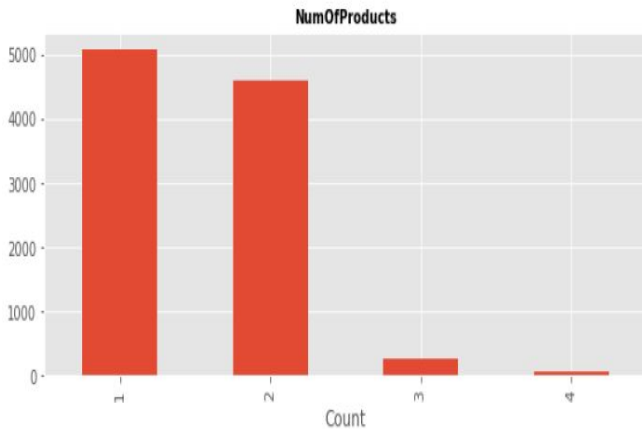
La base se conforma de 10,000 registros donde cada registro es la información de un cliente que pudo o no abandonar la relación con el banco.

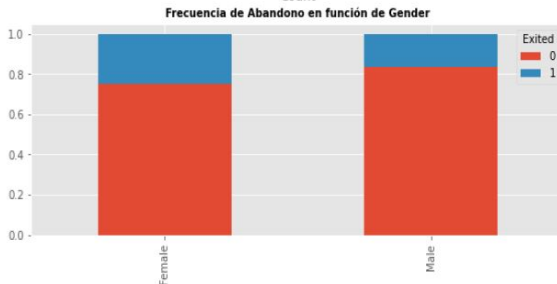
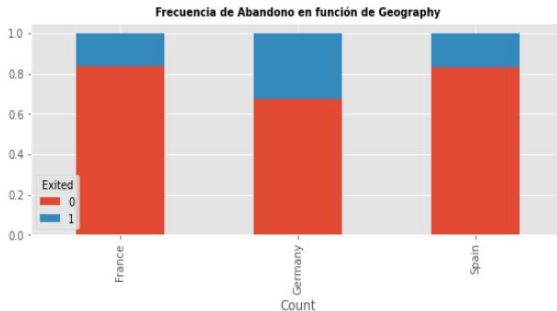
Features

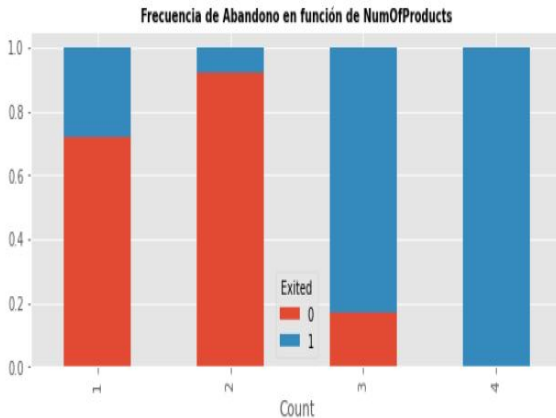
- CreditScore: Calificación del cliente, de tipo entero.
- Geography: Nacionalidad del cliente, de tipo object.
- Gender: Género del cliente, de tipo object.
- Age: Edad del cliente, de tipo entero.
- Tenure: Estado del cliente, de tipo entero.
- Balance: Balance en la cuenta del cliente, de tipo flotante.
- NumOfProducts: Número de productos que tiene el cliente con el banco, de tipo entero.
- HasCrCard: Si tiene asociada una tarjeta de crédito, de tipo entero.
- IsActiveMember: Si el cliente es activo, de tipo entero.
- EstimatedSalary: Salario estimado del cliente, de tipo flotante.
- Exited: Si el cliente abandonó al banco (1) o no (0) de tipo entero.

Análisis Exploratorio

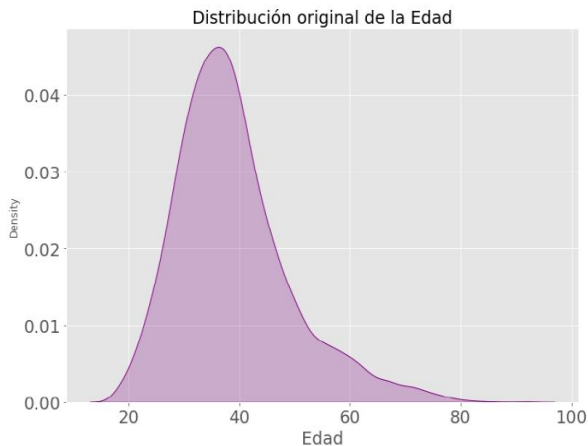




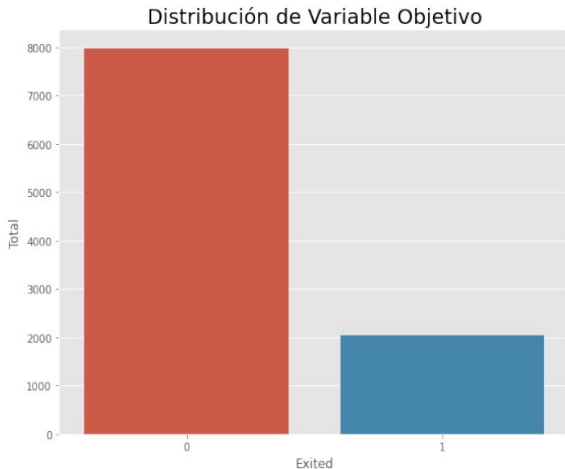




	count	mean	std	min	25%	50%	75%	max
CreditScore	10000.0	650.528800	96.653299	350.00	584.00	652.000	718.0000	850.00
Age	10000.0	38.921600	10.487806	18.00	32.00	37.000	44.0000	92.00
Tenure	10000.0	5.012800	2.892174	0.00	3.00	5.000	7.0000	10.00
Balance	10000.0	76485.889288	62397.405202	0.00	0.00	97198.540	127644.2400	250898.09
NumOfProducts	10000.0	1.530200	0.581654	1.00	1.00	1.000	2.0000	4.00
HasCrCard	10000.0	0.705500	0.455840	0.00	0.00	1.000	1.0000	1.00
IsActiveMember	10000.0	0.515100	0.499797	0.00	0.00	1.000	1.0000	1.00
EstimatedSalary	10000.0	100090.239881	57510.492818	11.58	51002.11	100193.915	149388.2475	199992.48
Exited	10000.0	0.203700	0.402769	0.00	0.00	0.000	0.0000	1.00

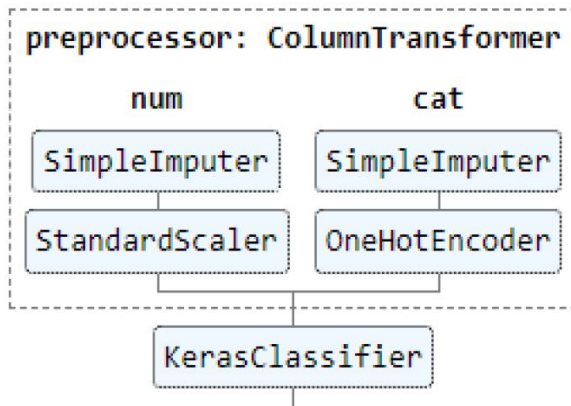


Valor	Total	Porcentaje
No Abandona	7963	79.63 %
Aprobada	2037	20.37 %

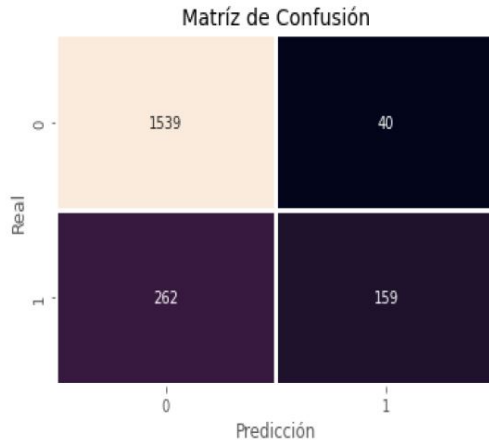


Modelo

Pipeline



Matriz de Confusión

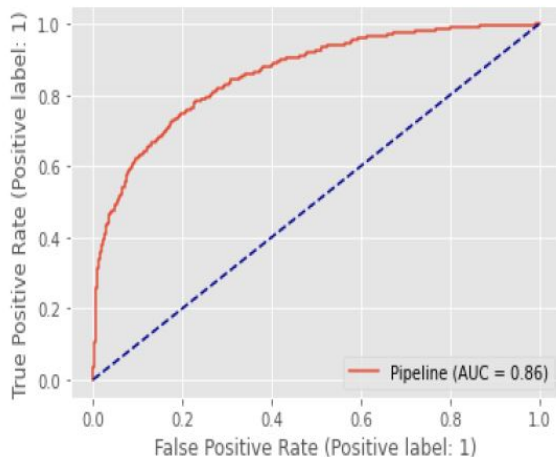


Métricas

Clase	Precision	Recall	f1-score
0 (No Abandona)	0.85	0.97	0.91
1 (Abandona)	0.80	0.38	0.51

Al aplicar una técnica de validación cruzada se obtiene un Accuracy promedio de 83.79 % con una desviación $\pm 0.79\%$

Curva ROC



Conclusiones

Conclusiones

- En este proyecto se consideró una dataset público conformado por clientes que decididieron abandonar un banco.

Conclusiones

- En este proyecto se consideró una dataset público conformado por clientes que decididieron abandonar un banco.
- Se planteó un problema de aprendizaje supervisado en donde la variable objetivo se representa con 1 si el cliente abandona 0 en caso contrario. Una vez limpia la base y al aplicar un análisis exploratorio de datos se aplicó un perceptrón multicapa.

Conclusiones

- En este proyecto se consideró una dataset público conformado por clientes que decidieron abandonar un banco.
- Se planteó un problema de aprendizaje supervisado en donde la variable objetivo se representa con 1 si el cliente abandona 0 en caso contrario. Una vez limpia la base y al aplicar un análisis exploratorio de datos se aplicó un perceptrón multicapa.
- Con respecto al modelo el 83.79% de los casos fueron correctamente clasificados utilizando un umbral de 0.5.

Conclusiones

- En este proyecto se consideró una dataset público conformado por clientes que decidieron abandonar un banco.
- Se planteó un problema de aprendizaje supervisado en donde la variable objetivo se representa con 1 si el cliente abandona 0 en caso contrario. Una vez limpia la base y al aplicar un análisis exploratorio de datos se aplicó un perceptrón multicapa.
- Con respecto al modelo el 83.79 % de los casos fueron correctamente clasificados utilizando un umbral de 0.5.
- También se calculó la curva ROC y se obtuvo un AUC de 0.86 que es bastante cercano a 1 por lo que se concluye que el modelo es capaz de hacer una buena clasificación.

Gracias