

# PROYECTO FINAL TERCERA ENTREGA

Julián Y. Torres Torres, Carlos J. Barreto Mora, Juan A. Angulo Rincón

Modelo de fuga de clientes

Universidad de los Andes, Bogotá, Colombia

{jy.torres, c.barretom, ja.angulor1}@uniandes.edu.co

<https://github.com/jaangulor/Ciencia-de-datos-aplicada---Proyecto-final.git>

Fecha de presentación: Diciembre 01 de 2022

<b>1. Introducción</b>	<b>1</b>
<b>2. Construcción y evaluación del modelo final</b>	<b>2</b>
<b>3. Despliegue del Modelo</b>	<b>3</b>
<b>4. Próximas actividades</b>	<b>3</b>

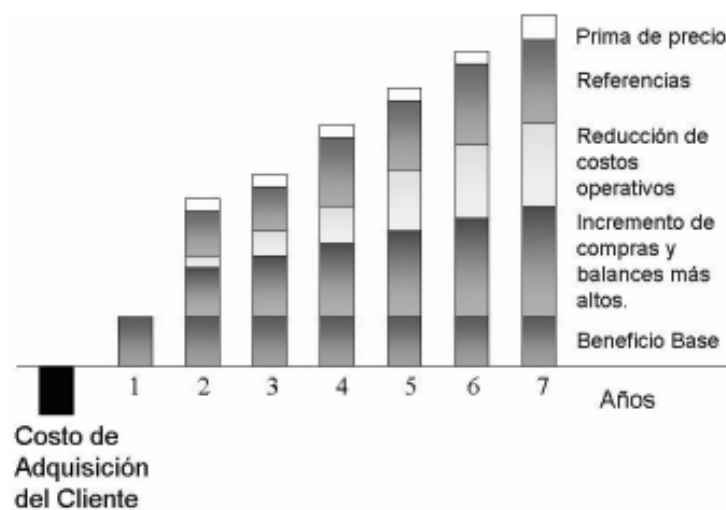
## 1. Introducción

En Colombia, las entidades financieras deben realizar actividades para la gestión del riesgo de crédito, de acuerdo a las directrices de la Superintendencia Financiera para mitigar la posibilidad que la entidad vigilada pueda incurrir en pérdidas y disminuir el valor de sus activos, cuando un deudor incumpla con sus obligaciones contractuales, así mismo se debe maximizar la tasa de rendimiento ajustada al riesgo del banco, manteniendo una exposición al riesgo de crédito dentro de los parámetros aceptables, de acuerdo a lo establecido en Basel Committee on Banking Supervision.

Por esa razón la entidad financiera que se seleccionó y mantendremos su nombre bajo reserva , debe realizar políticas de otorgamiento de cartera, desarrollo de modelos para medir la probabilidad de incumplimiento de algunos de sus clientes, definir cuáles clientes son sujetos de créditos, manejo eficiente del apetito de riesgo y riesgo - retorno. Debido a todo lo anterior, actualmente las entidades financieras se encuentran muy interesadas en el uso de la Inteligencia Artificial, Aprendizaje Automático (Machine Learning), diseñando estrategias de transformación digital, mejora de procesos, reducción de costos y detección eficiente de los fraudes, también todo lo correspondiente a la fidelización de los clientes.

## Problema del Negocio :

Debido a la alta competitividad existente en el sector financiero , las entidades financieras deben realizar un mejor manejo de sus recursos, y ofrecer nuevos productos financieros a sus clientes actuales y potenciales. También se debe realizar mediciones a la satisfacción o insatisfacción del cliente con el servicio prestado, evitando así la fuga de clientes a la competencia. Las entidades bancarias que poseen una cartera con un mayor número de clientes tiende a tener utilidades más altas en la institución debido a la alta transaccionalidad, menores gastos operacionales y en especial y no menos importante, se traduce en la confianza y fidelidad del cliente, que a través del tiempo produce menos costos en la organización, pudiendo referenciar a la entidad con más personas.



**Figura 1:** Beneficio monetario que otorga un cliente a lo largo del tiempo y que permanece en la institución [1].

De acuerdo a la necesidad de negocio de la organización a la cual se le hará el análisis de negocio, se definió la fuga de clientes para esta entidad financiera como **“clientes que abrieron productos de colocación que deciden por iniciativa propia poner fin a sus obligaciones con la entidad o clientes que no han evidenciado actividad en sus obligaciones con la entidad en los últimos 6 meses”**.

La solución propuesta a las diferentes problemáticas anteriormente mencionadas, plantean a la construcción de diversos modelos de clasificación binaria, el cual asigna a cada cliente una de las dos clases predefinidas (F/NF) basándonos en un conjunto de variables de entrada previamente identificadas y definidas que pueden incidir en el negocio.

## Aproximación desde la analítica de datos :

Actualmente la entidad bancaria que cuenta con la necesidad de implementar políticas de fidelización y retención, no cuenta con estrategias que permitan comparar con las soluciones obtenidas en este proyecto, es decir; es el primer planteamiento de modelo de fuga que se hace en la institución a través de métodos machine learning.

## Metodología :

En la aplicación de los diferentes modelos de machine learning es imprescindible diseñar una cadena de procesos , que garantice una buena calidad de los datos que proviene de la entidad financiera, realizar todas las actividades de limpieza de datos, y posteriormente modelación y evaluación de cada uno de los modelos.



**Figura 2 :** Machine Learning para la estimación del riesgo de crédito en una cartera de consumo [2]

Finalmente, la implementación del modelo de fuga deber ser un preámbulo a una secuencia de acciones que permitan retener clientes identificando los clientes que tengan tendencias más altas de fuga y también determinar cuál es la rentabilidad que se puede esperar de estos clientes para así, focalizar la atención de la organización en el mejor potencial.

A continuación podemos visualizar un esquema en el cual podemos entender el ciclo de la retención, teniendo en cuenta que se realizó una correcta implementación de un modelo predictivo de fuga



**Figura 3:** Ciclo de retención de clientes

Un ciclo de retención de clientes estratégico y diseñado de manera adecuada, promueve gestiones comerciales fundamentales, tales como la captación y retención de clientes que permiten aumentar el tamaño de la cartera e idear estrategias más focalizadas disminuyendo los gastos de inversión en la búsqueda de acaparar nuevos clientes del mercado (La figura 1 evidencia el costo de vincular clientes nuevos en la organización).

## 2. Construcción y evaluación del modelo final

La base de datos cuenta con un potencial de 1 '050.742 clientes segmentados en la banca agropecuaria de una entidad financiera. Sin embargo como plan estratégico, se tomaron las regionales de Bogotá, Santander, Sur y Cafetera que son las regionales en donde se evidenció que la entidad financiera tenía mayor cantidad de clientes “Fugados” y no “Fugados”, finalmente trabajando con un potencial de 509297 clientes persona natural.

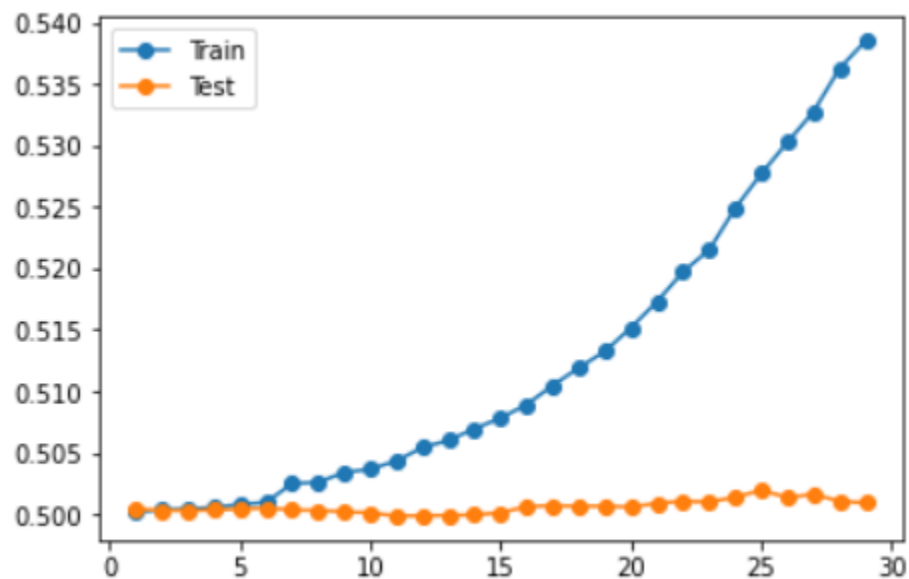
La base de datos fue debidamente trabajada para la extracción de datos atípicos y valores nulos que pudiesen interferir en el análisis posteriormente realizado. También se tuvieron en cuenta únicamente clientes como único registro, es decir sin tratamiento de duplicados para no sesgar de alguna manera las predicciones del modelo.

A continuación podremos evidenciar los resultados obtenidos de nuestro mejor modelo:

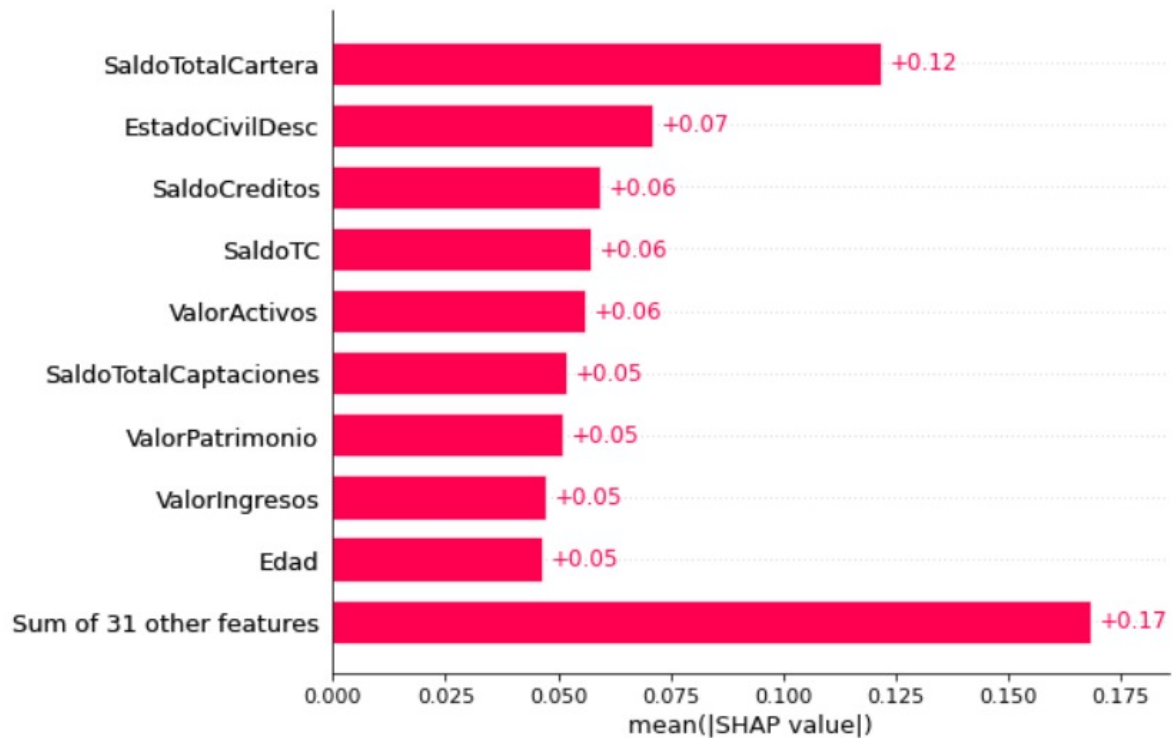
### Evaluación cuantitativa :

Modelo	Exactitud
Árbol de decisión	0.500
XGBoost	0.499
RadomForest	0.499
Logistic	0.499
GNB	0.500
Baggin	0.489
Redes neuronales	0.501
GMB	0.513

### Variación de hiper parámetros:



Para la validación del modelo procedimos a realizar validación cruzada y variación de parámetros, donde evidenciamos que al aumentar la profundidad de árbol de decisión aumenta la cantidad de overfitting.



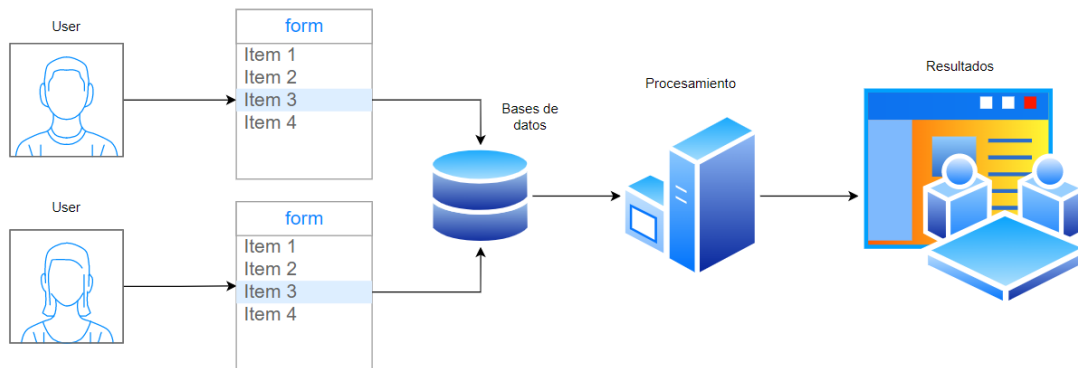
Evaluación cualitativa :

Todos los modelos presentaron resultados positivos de acuerdo a la evaluación realizada en cada uno de ellos. El modelo Árbol de decisión con la métrica analítica de 0.5 , permitirá al banco realizar actividades para la identificación de los clientes que se encuentran insatisfechos con la entidad. Si la entidad financiera busca alcanzar una probabilidad de churn del 70% sobre los clientes, deberá realizar la implementación de este modelo en una etapa productiva para encontrar los clientes cuya probabilidad se encuentre por debajo del 70%. De esa forma la entidad podrá ejecutar acciones de forma proactiva para tener acercamiento con esos clientes y brindarles una mejor experiencia , para cambiar la percepción que tienen hacia la entidad financiera.

Cuando se encuentre desplegado en producción se recomienda realizar monitoreo y actualizar el modelo una vez al mes con el fin de hallar cualquier modificación en el comportamiento de los clientes.

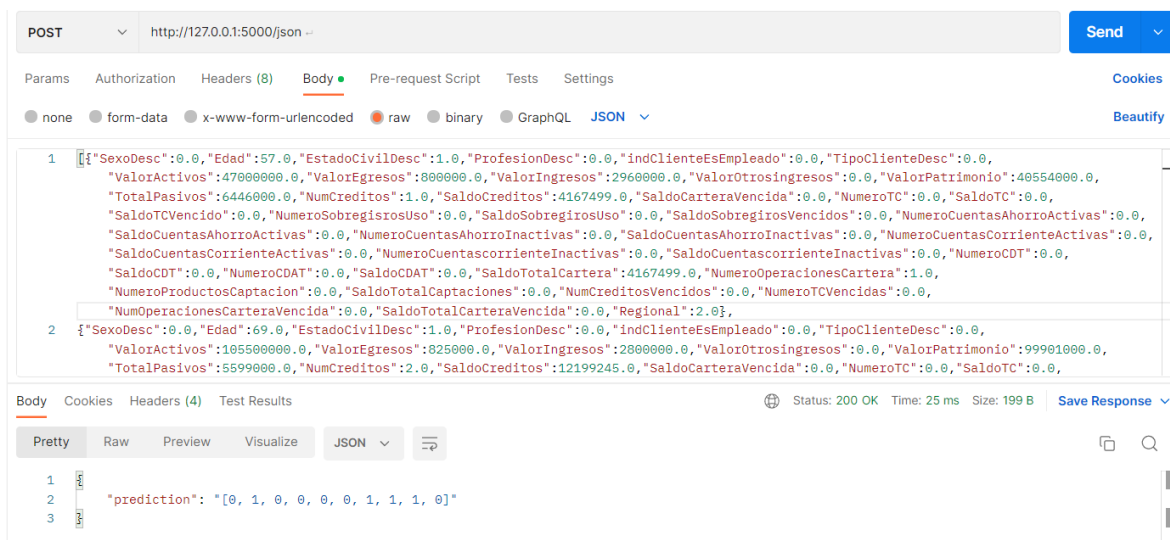
### 3.Despliegue del Modelo

Para el despliegue del modelo se propone la arquitectura que se muestra en la figura a continuación.



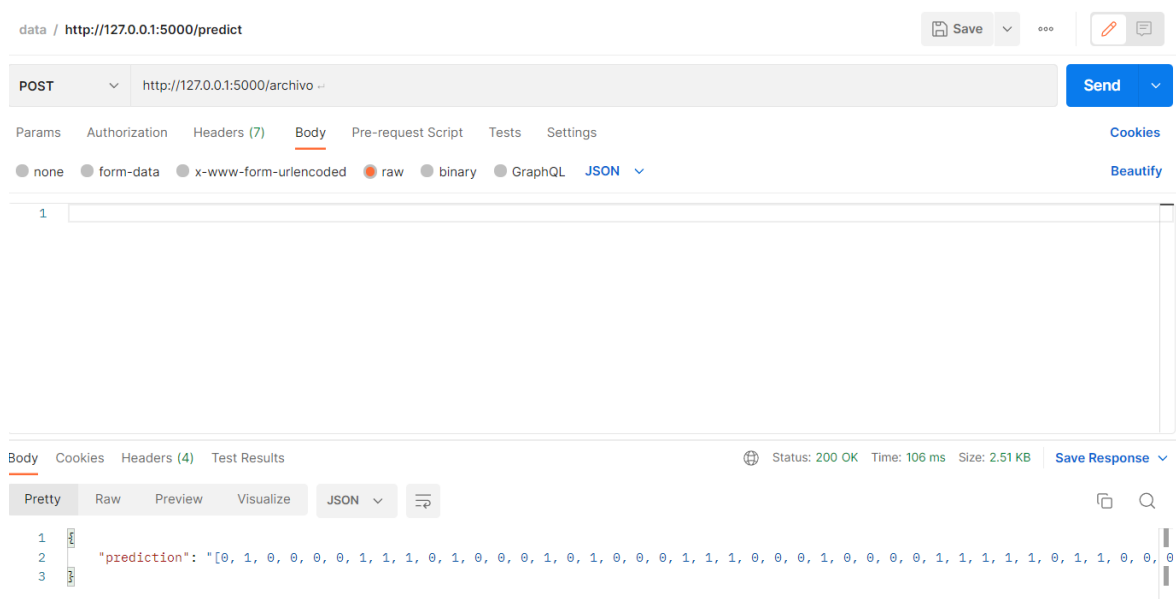
**Figura 4.** Esquema general de arquitectura propuesta para el proyecto.

Consiste en bases de datos de clientes existentes, para los que es necesario determinar la fuga. Estos datos se procesan por medio del API (Python enlazado con Flask) y se muestran al cliente los resultados de fuga mediante postman. Se realizan dos rutas posibles para el procesamiento, la primera llamada ‘/json’ permite el procesamiento de uno o varios clientes, pero para ello es necesario adicionar los datos en el body-row en formato json en postman, como se muestra en la siguiente figura.



**Figura 5.** Procesamiento de un número determinado de clientes con ingreso manual.

Adicionalmente es posible realizar el procesamiento por lotes generando un archivo tipo json que el API lee y procesa directamente. Esto es muy útil para cuando se requiera hacer la evaluación de múltiples clientes al mismo tiempo.



**Figura 6.** Procesamiento con una base de datos.

## 4. Conclusiones

En este proyecto se logró poner en práctica las diferentes técnicas aprendidas en el curso de Ciencia de Datos Aplicada. Se aplicaron estrategias para problemas de alta dimensionalidad, desbalance de los datos . Trabajar con una base de datos con variables explicativas y un target objetivo (clientes insatisfechos ). Adicionalmente se logró obtener una métrica de machine learning por encima de 0.5, con correlaciones entre las variables explicativas y el target .

Nos permitió realizar una etapa de preprocesamiento riguroso , calidad de los datos, en donde aplicamos una variedad de estrategias diferentes para hacer la limpieza, depuración, y transformación de características que lograron reducir la dimensionalidad de los datos .

Posteriormente a la etapa de preprocesamiento , la búsqueda de hiperparámetros por cada modelo fue exitoso, obteniendo resultados de cada modelo por encima de un 0.5

De acuerdo con lo anterior la entidad financiera puede diseñar políticas para los clientes con focos de fidelización. Esta propuesta desarrollada por nosotros hacia la entidad financiera busca la aplicación del modelo a sus clientes



objetivos de negocio para no perderlos o evitar que se vayan con la competencia. Cuando se identifiquen clientes insatisfechos la entidad pueda realizar actividades como : brindar asesoría personalizada, brindar beneficios económicos, programa de fidelización, entre otras que permitan ofrecer una mejor experiencia hacia el consumidor financiero.

## 5. Bibliografía

1. F. Reichheld and E. Sasser. Zero defections: Quality comes to services. Harvard Business Review, 1990:105–111, September–October, 2000.
2. Machine Learning para la estimación del riesgo de crédito en una cartera de consumo. Universidad EAFIT, Maestría en Administración Financiera , 2021
3. L.M. Hitt and F.X. Frei. Do better customers utilize electronic distribution channels? the case of pc banking. Management Science, 48(6):732–748, 2002.
4. C.B. Bhattacharya. When customers are members: Customer retention in paid membership context. Journal of the Academy of Marketing Science, 26(1):31–44, 1998.
5. J. Dawes and S. Swailes. Retention sans frontieres: issues for financial service retailers. International Journal of Bank Marketing, 17(1):36–43, 1999.