

- INSTITUTO SUPERIOR POLITÉCNICO CÓRDOBA –
TECNICATURA EN CIENCIAS DE DATOS E INTELIGENCIA ARTIFICIAL

PROFESOR: NARCISO PEREZ

INTEGRANTES: CHAYLE, CHRISTIAN - MEDINA, JORGE

PROYECTO FINAL - RED NEURONAL (aplicar las redes neuronales a la predicción y descripción del siguiente problema)

PROYECTO: Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos.

PRESENTACIÓN

En la actualidad, el rol que cumplen los bancos en la economía de un país y el impacto que tienen en las diferentes clases sociales es cada vez más importante.

Estos han sido un mercado que históricamente ha recibido un gran número de quejas y reclamos. Es por ello por lo que, un mal servicio por parte del proveedor, una deficiente calidad de los productos, son las principales razones por las que los clientes abandonan una entidad bancaria.

Este inconveniente va en aumento y los bancos muestran su preocupación por este problema intentando implementar modelos que hasta el momento no han logrado cumplir con los objetivos.

Además, existe un elevado nivel de competencia que obliga a las entidades a velar por la lealtad de sus clientes para intentar mantenerlos e incrementar su rentabilidad.

En este proyecto proponemos realizar un análisis predictivo e implementar una Red Neuronal, con el fin de ayudar en la toma de decisiones para la retención de los clientes potenciales con tendencia a la deserción en la entidad bancaria. Esta propuesta se realizó mediante la necesidad de los mismos de conocer la exactitud de deserción de sus clientes categorizados potencialmente.

Se desarrolló un modelo predictivo mediante Redes Neuronales, con el fin de mostrar el resultado obtenido por el modelo indicando la exactitud, en porcentaje, de los clientes con tendencia a desertar.

INTRODUCCIÓN

El desarrollo de este proyecto tiene como objetivo principal implementar un Modelo de Análisis Predictivo que permita determinar la exactitud de los clientes con tendencias a la deserción en la provincia de Córdoba.

OBJETIVO

Este proyecto de investigación propone un Modelo de Análisis Predictivo. Se realizó la recolección de información real de clientes de un banco para que pueda ser trabajado, analizado e insertado en el modelo propuesto.

El desarrollo de este proyecto tiene como objetivo principal implementar un Modelo de Análisis Predictivo que permita determinar clientes con tendencias a la deserción en los bancos para apoyar la toma de decisiones de la entidad financiera que lo utilice.

PLANTEAMIENTO DEL PROBLEMA

La deserción de los clientes bancarios aumenta cada año en el país, provocando pérdidas monetarias millonarias, esta tendencia está directamente relacionada con un mal servicio, la baja calidad de los productos y la alta competencia entre las entidades es una necesidad prioritaria de los bancos contar con información y herramientas precisas que apoyen la toma de decisiones para la implementación de acciones pertinentes para retener a los clientes potenciales con tendencias a la deserción.

Por tal motivo, es importante analizar el comportamiento de los clientes que previamente han desertado en el banco, para así tener un estándar base que se pueda comparar a los clientes actuales que tiene la entidad bancaria.

Los resultados que se obtienen ayudan a los bancos a tomar planes de acción para poder conservar a sus clientes potenciales ya que son aquellos clientes que invierten más en la entidad bancaria.

Apoyar la toma de decisiones para retener clientes potenciales con tendencias a la deserción en las distintas entidades bancarias.

Causas:

- Los modelos actuales cubren otros aspectos sin foco en el cliente, como, por ejemplo, la venta excesiva de sus productos sin existir alguna necesidad.

Es por ello que se requiere desarrollar una alternativa enfocada a recuperar a los clientes potenciales que tienden a desertar.

- Los Modelos actuales trabajan con pequeñas cantidades de información y es por ello que se requiere de una herramienta aplicada a un modelo que permita al análisis de grandes cantidades de datos.

RESOLUCIÓN DEL PROBLEMA

Con este modelo se busca identificar que clientes tienen mayor probabilidad de desertar del banco en donde se encuentran registrados. Para ello, se analiza el comportamiento de los clientes que previamente han desertado en el banco, para así

tener un estándar base que se pueda comparar a los clientes actuales que tiene la entidad bancaria.

Los resultados que se obtienen ayudan a los bancos a tomar planes de acción para poder conservar a sus clientes potenciales ya que son aquellos clientes que invierten más en la entidad bancaria.

OBJETIVOS DEL PROYECTO

- Objetivo General

Implementar un Modelo de Análisis Predictivo que permita determinar clientes con tendencias a la deserción en los bancos de la ciudad de Córdoba.

- Objetivos Específicos

* Analizar las herramientas de Machine Learning que permitan la implementación de un modelo de análisis predictivo

de clientes desertores enfocado a las instituciones bancarias.

* Diseñar un Modelo de Análisis Predictivo que permita pronosticar los clientes con tendencias a la deserción.

* Elaborar un plan de continuidad del modelo que permita pronosticar clientes desertores de las instituciones bancarias.

PLANIFICACIÓN DEL PROYECTO

El proyecto se desarrolla en base al análisis de la información del comportamiento de los clientes desertores de las entidades bancarias, trabajando con gran cantidad de datos que se ingresan a la herramienta para diseñar el Modelo de Análisis predictivo.

MARCO TEÓRICO

Se realiza la descripción de las definiciones de todos los términos y conceptos esenciales en el ámbito de predicción para permitir que el lector se familiarice

con el tema del proyecto. Dentro de los conceptos a desarrollar, se encuentran conceptos básicos, complementarios y específicos al contexto de análisis predictivo.

Datos: Es un registro que caracteriza un hecho de la realidad. Consiste en una representación simbólica (alfabética, numérica, gráfica, etc.) de una variable o atributo cuantitativo o cualitativo.

El dato por sí solo, sin ser contextualizado, no es valioso. Su valor deviene de la posibilidad de transformarlo en información útil para la toma de decisiones.

Información: La información es un mensaje que tiene un emisor y un receptor. Surge del procesamiento de datos, que son resumidos y analizados, transformados con la intencionalidad y objetivo de comunicar algo.

¿QUÉ ES EL ANÁLISIS PREDICTIVO?

Los Análisis Predictivos agrupan una variedad de técnicas estadísticas de modelización, aprendizaje automático y minería de datos y por ello siempre es necesario revisar su desempeño y poder de predicción para ser mejorados o inclusive cambiado en sus parámetros ya que los patrones de comportamiento pueden variar según la época o momento en el que fueron desarrollados.

Dentro de las empresas se genera información estructurada para el desarrollo de modelos predictivos proveniente de los sistemas transaccionales, así como datos no estructurados que vienen de otras áreas como servicio al cliente para conocer la opinión acerca de productos brindados por las entidades bancarias.

APRENDIZAJE SUPERVISADO Y MACHINE LEARNING

El objetivo de un algoritmo de aprendizaje supervisado es obtener una clasificación a través de lo aprendido de ejemplos de entrenamiento.

Esta clasificación es algo que puede ser usado para hacer predicciones en ejemplos de pruebas.

Dentro del contexto de aprendizaje supervisado nace el término de Machine Learning o aprendizaje de máquina, el cual hace referencia a la detección

de manera automática de patrones significantes en datos. Una característica de la aplicación de machine learning es que se aplica en casos

donde los patrones a identificar son de alta complejidad, por lo que el ser humano no cuenta con la pericia suficiente para aprender de esta manera específica.

MODELOS PREDICTIVOS

Es un nombre dado a una colección de técnicas matemáticas con la meta en común de encontrar una relación matemática entre un objetivo,

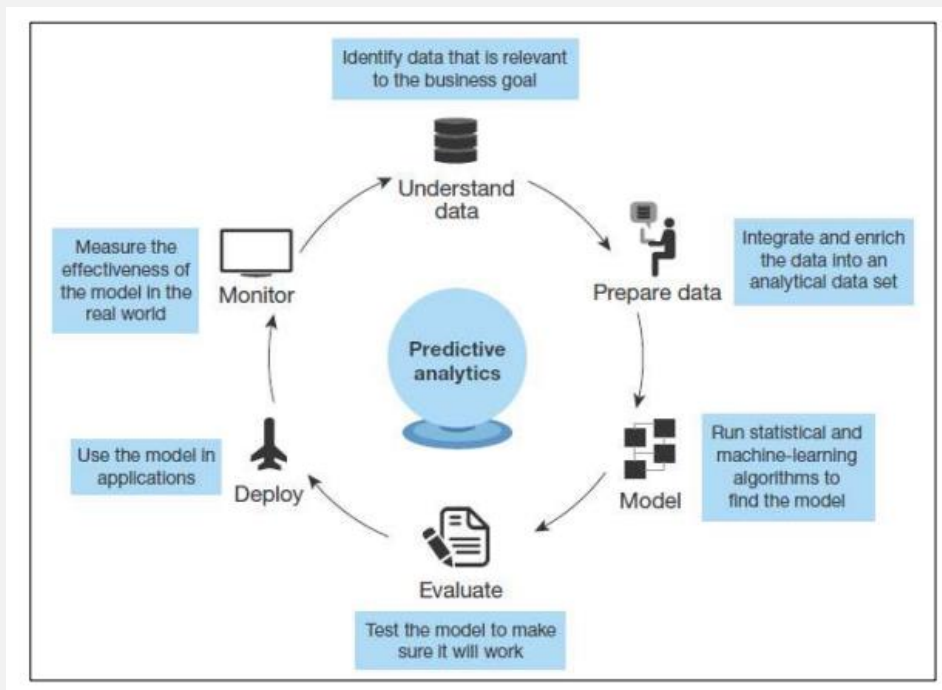
respuesta o variable dependiente y varios factores de predicción y variables independientes, con el objetivo de medir valores futuros

de estos factores de predicción e insertándolos en la relación matemática para predecir valores futuros de la variable objetivo.

Como estas relaciones nunca son perfectas en la práctica, es deseable dar una solución de duda en las predicciones.

Para identificar aquellos factores que interceden e la predicción, se pueden agrupar en tres categorías: aquellos que tienen poca posibilidad de afectar el resultado, aquellos con cierta certeza para afectar los resultados y deben ser considerados en el modelo y los que se encuentran en el medio, que pueden influir o no en el resultado final, siendo necesario identificar a través de una serie de técnicas si deben ser incluidas en el modelo.

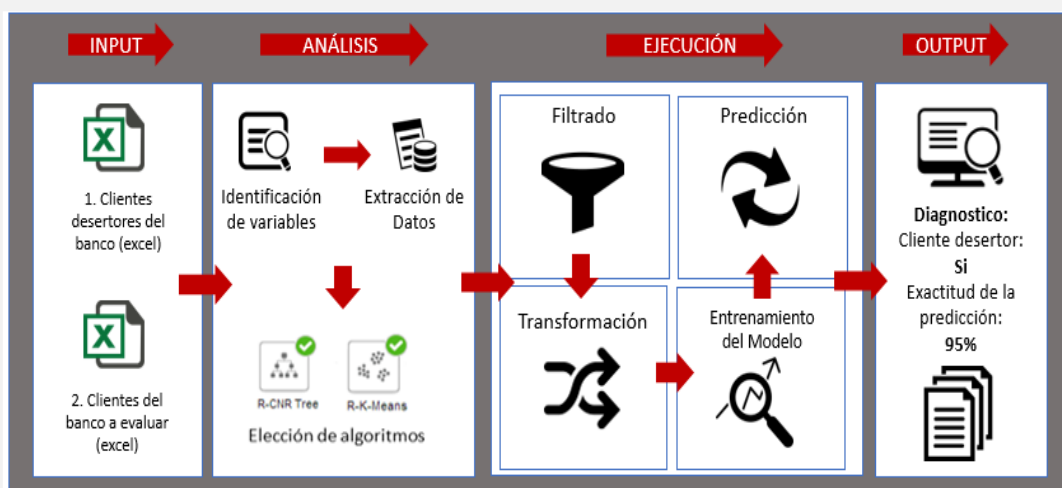
6 PASOS DEL ANÁLISIS PREDICTIVO



METODOLOGÍA

Elaboración del Modelo de Análisis Predictivo

Este modelo propuesto permite la gestión temprana de retención y fidelización de los clientes con tendencia a la deserción de una entidad bancaria. Se identificaron tres factores importantes para el diseño de este modelo: la carga y procesamiento de datos a través de un archivo .csv con datos de clientes de dicha entidad bancaria, la selección de variables, algoritmos y ejecución del modelo.



El modelo está compuesto por lo siguiente:

La carga y procesamiento de datos se realiza a través de la carga de archivo Excel (csv), conteniendo información de los clientes de la entidad bancaria. El archivo contiene información del comportamiento de los clientes que se utilizó para analizar el modelo.

Cientes_banco.csv

Se importó el dataset en un entorno IDE (Entorno de Desarrollo Integrado) local de Python en Jupyter Lab de Anaconda Navigator.

IMPORTACIÓN DE LIBRERIAS Y PROCESAMIENTOS DE DATOS

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

# IMPORTANDO DATASET.

dataset_csv = pd.read_csv('Cientes_banco.csv')

# COLUMNAS DEL DATASET.

print("\nColumnas del DataSet: ")
print("=====")
print(dataset_csv.columns)
print("=====")

# DESCRIBIR EL DATASET ORIGINAL.

print("\nDataset original:\n", dataset_csv.describe(include='all'))
print("=====")

# DATASET REDUCIDO.

dataset = dataset_csv.iloc[:,3:14]
dataset_columns = dataset.columns
dataset_values = dataset.values

# DESCRIBIR EL DATASET TRUNCADO.

print("\nDataset reducido: ")
print("\n",dataset.head())
print("=====")
# Revisamos los tipos de datos de las Columnas
print("\nTipos de Columnas del Dataset: ")
print(dataset.dtypes)
print("=====")
```

El notebook que se adjunta al proyecto se estructuró en:

- ✓ Importación de Dataset de Features Clientes y su valoración.
- ✓ Análisis Descriptivo Exploratorio.
- ✓ Desarrollo de Modelo de Predicción de Aprendizaje Profundo (Deep Learning).

Análisis Exploratorios de Datos

Luego de realizar las principales validaciones del dataset importado con los comandos básicos de python, con librerías como Pandas, tales como `dataset_csv= pd.read_csv('Cientes_banco.csv')`, la primera fase es el proceso de filtrado, con el objetivo de limpiar los datos o rellenar campos nulos según corresponda, verificar tipos de datos y eliminar duplicados en caso de existir, con la finalidad de tener un conjunto de datos que puedan ser analizados correctamente, comenzamos con un primer análisis descriptivo de ciertas variables elegidas del dataset para analizar distintos ejes del dataset.

```
In [16]: dataset_csv= pd.read_csv('Clientes_banco.csv')
dataset_csv
```

```
Out[16]:
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	
	0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
	1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
	2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
	3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0
	4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

	9995	9996	15606229	Obijaku	771	France	Male	39	5	0.00	2	1	0	96270.64	0
	9996	9997	15569892	Johnstone	516	France	Male	35	10	57369.61	1	1	1	101699.77	0
	9997	9998	15584532	Liu	709	France	Female	36	7	0.00	1	0	1	42085.58	1
	9998	9999	15682355	Sabbatini	772	Germany	Male	42	3	75075.31	2	1	0	92888.52	1
	9999	10000	15628319	Walker	792	France	Female	28	4	130142.79	1	1	0	38190.78	0

10000 rows × 14 columns

Columnas del DataSet:

=====

```
Index(['RowNumber', 'CustomerId', 'Surname', 'CreditScore', 'Geography',
      'Género', 'Edad', 'Tenencia', 'Saldo', 'Número de productos', 'HasCrCard',
      'IsActiveMember', 'EstimatedSalary', 'Exited'],
      dtype='objeto')
=====
```

Conjunto de datos original:

```
RowNumber CustomerId Apellido CreditScore Geografía Género \
cuenta 10000.00000 1.000000e+04 10000 10000.000000 10000 10000
único NaN NaN NaN 2932 NaN 3 2
arriba NaN NaN Smith NaN Francia Hombre
frecuencia NaN NaN 32 NaN 5014 5457
media 5000.50000 1.569094e+07 NaN 650.528800 NaN NaN
estándar 2886.89568 7.193619e+04 NaN 96.653299 NaN NaN
min 1.00000 1.556570e+07 NaN 350.000000 NaN NaN
25% 2500.75000 1.562853e+07 NaN 584.000000 NaN NaN
50% 5000.50000 1.569074e+07 NaN 652.000000 NaN NaN
75% 7500.25000 1.575323e+07 NaN 718.000000 NaN NaN
máx. 10000.00000 1.581569e+07 NaN 850.000000 NaN NaN
```

```
Edad Tenencia Saldo NumOfProducts HasCrCard \
cuenta 10000.000000 10000.000000 10000.000000 10000.000000 10000.000000
único NaN NaN NaN NaN NaN NaN
superior NaN NaN NaN NaN NaN NaN
frecuencia NaN NaN NaN NaN NaN NaN
media 38.921800 5.012800 76485.889288 1.530200 0.70550
estándar 10.487806 2.892174 62397.405202 0.581654 0.45584
min 18.000000 0.000000 0.000000 1.000000 0.000000
25% 32.000000 3.000000 0.000000 1.000000 0.000000
50% 37.000000 5.000000 97198.540000 1.000000 1.000000
75% 44.000000 7.000000 127644.240000 2.000000 1.000000
máx. 92.000000 10.000000 250898.090000 4.000000 1.000000
```

```
Salario estimado de IsActiveMember Exited
cuenta 10000.000000 10000.000000 10000.000000
único NaN NaN NaN
superior NaN NaN NaN NaN
frecuencia NaN NaN NaN NaN
media 0.515100 100090.239881 0.203700
estándar 0.499797 57510.492818 0.402769
min 0.000000 11.580000 0.000000
25% 0.000000 51002.110000 0.000000
50% 1.000000 100193.915000 0.000000
75% 1.000000 149388.247500 0.000000
máx. 1.000000 199992.480000 1.000000
=====
```



```

Conjunto de datos reducido:

    CreditScore Geografía Género Edad Tenencia Balance NumOfProducts \
0 619 Francia Mujer 42 2 0,00 1
1 608 España Mujer 41 1 83807.86 1
2 502 Francia Mujer 42 8 159660.80 3
3 699 Francia Mujer 39 1 0,00 2
4 850 España Mujer 43 2 125510,82 1

    HasCrCard IsActiveMember CalculatedSalary Exited
0 1 1 101348.88 1
1 0 1 112542.58 0
2 1 0 113931.57 1
3 0 0 93826.63 0
4 1 1 79084.10 0
=====

Tipos de Columnas del Dataset:
Puntaje crediticio int64
Objeto de geografía
Objeto de género
Edad int64
Tenencia int64
Equilibrio float64
NumOfProducts int64
HasCrCard int64
Es un miembro activo int64
Flotación de salario estimado64
Salió int64
tipo: objeto
=====

```

Esto nos permite conocer la distribución de las principales variables del dataset, por ejemplo: en este caso tenemos variables categóricas en las columnas País y Genero.

Análisis de Correlación

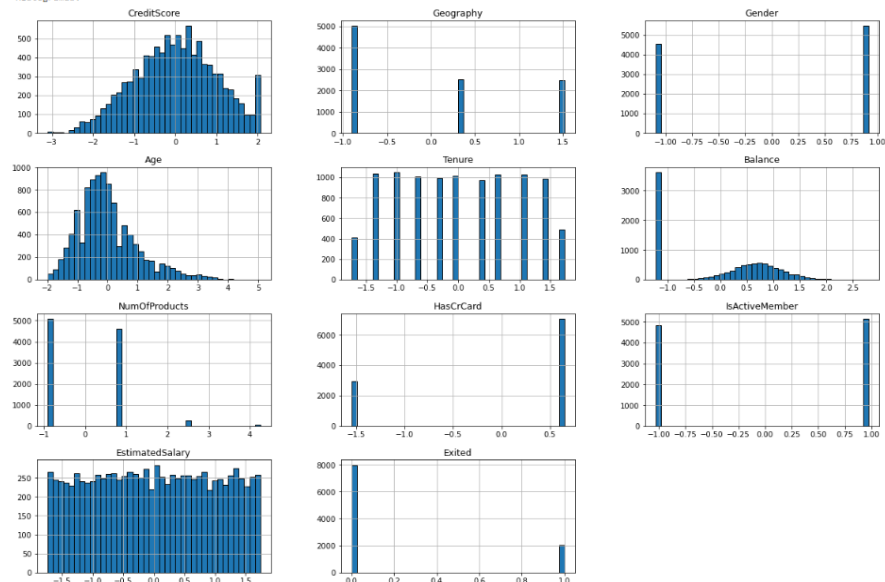
A continuación, se realiza un análisis de correlación de variables para entender si podemos detectar visualmente ciertas relaciones claras entre las variables del dataset.

```
[6]: # DISTRIBUCIONES DEL DATA SET Y CORRELACIONES.
```

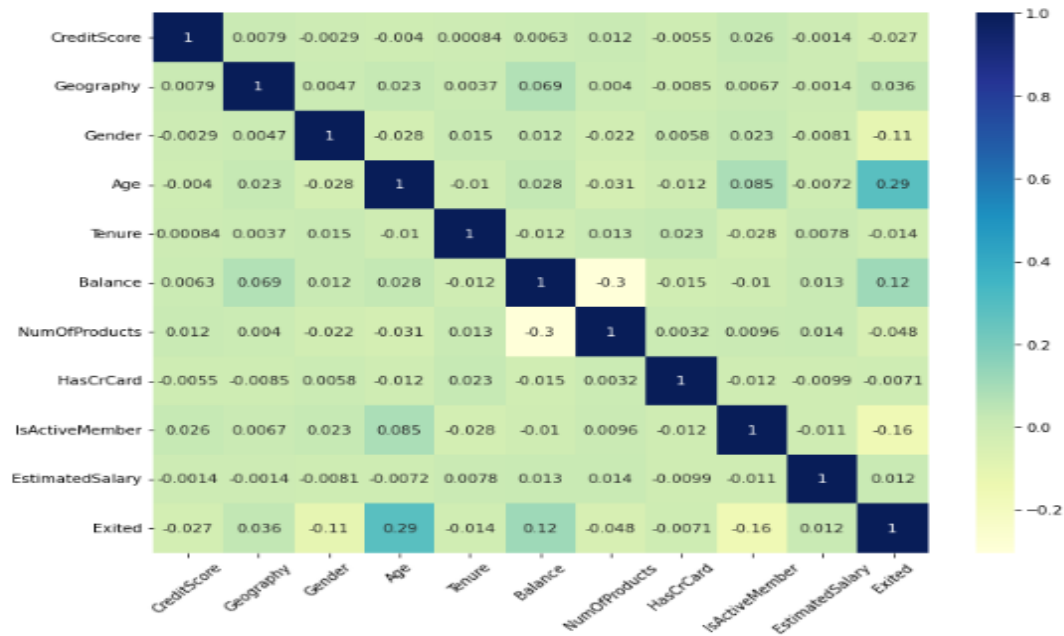
```
print("\n Histogramas:")
plotHistogram(dataset_final)

print("\n Correlaciones:")
plotCorrelations(dataset_final)
```

Histogramas:



Correlaciones:



Podemos observar en el gráfico que las variables se encuentran correlacionadas.

EXPERIMENTOS Y ANALISIS

✓ Instancias de Pruebas

El presente proyecto propuesto es trabajado con un total de 2.000 clientes, siendo 1548 clasificados como no desertores y un total de 141 clientes clasificados como desertores.

✓ Evaluamos el modelo y calculamos las métricas de evaluación del mismo.

```
[14]: from sklearn.metrics import classification_report
      print(classification_report(y_test, y_pred_norm))
```

	precision	recall	f1-score	support
0	0.85	0.97	0.91	1595
1	0.75	0.35	0.48	405
accuracy			0.84	2000
macro avg	0.80	0.66	0.69	2000
weighted avg	0.83	0.84	0.82	2000

Para evaluar correctamente un modelo de predicción con clases desbalanceadas, nos interesa no sólo obtener un buen resultado en el Accuracy "Exactitud", es decir el porcentaje de caso que el modelo ha acertado "verdaderos positivos + verdaderos negativos / "verdaderos positivos + verdaderos negativos + falsos positivos + falsos negativos", sino fundamentalmente obtener un buen rendimiento respecto de las métricas Precisión "precisión", definida como verdaderos positivos / "verdaderos positivos + falsos positivos", y Recall "exhaustividad", definida como verdaderos positivos / "verdaderos positivos + falsos negativos". La métrica Precisión nos informará cuán bueno es el modelo en identificar la clase correcta, mientras que la métrica Recall nos ayudará a entender cuanto identifica la clase correcta.

CONCLUSIONES

En esta sección se destacan las conclusiones principales obtenidas en este trabajo de investigación y se resumen los resultados obtenidos del modelo propuesto. Se comentan algunos aspectos relacionados con los trabajos futuros que siguen la propuesta planteada y sobre otros temas de investigación que se pueden derivar. El principal propósito de la elaboración de un Modelo de Análisis Predictivo se basa en la aplicabilidad en un entorno organizacional que no genere el estancamiento de las actividades de la empresa. El planteamiento del modelo consiste en que una entidad bancaria pueda fidelizar y retener a sus clientes potenciales. De esta manera, con la tecnología actual del mercado se propuso un Modelo de Análisis Predictivo para la retención de los clientes con tendencias a la deserción en entidades bancarias, a través del comportamiento de los clientes en cuanto a sus transacciones, movimientos y actividad financiera. El modelo diseñado cumple con el propósito de pronosticar clientes con tendencias a la deserción en Bancos a través del análisis del comportamiento, alcanzando la mejor precisión (0.8627) con el algoritmo de redes neuronales.

El Modelo de Análisis Predictivo utilizó datos recopilados para el proyecto, siendo clave resaltar que, con el uso de una mayor cantidad de datos, y con la característica que sean de origen bancario, se podrían mejorar los resultados obtenidos. Los resultados experimentales de la investigación del proyecto muestran que el modelo puede alcanzar una buena precisión-predicción de clientes desertores.