

**Calculo de probabilidad de que Clientes bancarios realicen Depósitos a Plazo**

Integrantes del proyecto: Juan Cassinerio - Juan Isasi - Jose Vargas

Fecha de actualización: 18/03/2023

**Contenido:**

* 1. Introducción
  2. Dataset y Análisis de Variables

2.1 Objetivos

2.1.1 **Relación entre estado civil del cliente y Probabilidad de realizar depósitos a plazo**

2.1.2 **Meses en que los clientes son más predispuestos a realizar depósitos a plazo**

2.1.3 **Relación entre edad del cliente y duración del último contacto con el mismo con la Probabilidad que este efectué un depósito a plazo**

* 1. Desarrollo del Modelo

3.1 **Modelo por Regresión Lineal**

3.2 **Modelo por Decision Tree**

3.3 **Modelo por Random Forest**

* 1. Conlusiones

1. **Introducción**

* ***Contexto empresarial***

Dentro de los productos y servicios que ofrecen los bancos, los depósitos a plazo representan un gran porcentaje de las fuentes de ingresos. Para poder maximizar la cantidad de depósitos a plazo realizados, se buscará identificar a los clientes (parámetros) que tengan la mayor probabilidad de subscribirse a este tipo de producto y de esa manera poder mejorar los esfuerzos de marketing en dichos clientes.

* ***Problema empresarial***

Se recopilo un conjuntos de datos con el objetivo de desarrollar modelos de **machine learning** para predecir la probabilidad de que clientes realicen depósitos a plazo en bancos. Estos datos fueron tomados de una campaña de marketing de una importante institución bancaria de Portugal.

* ***Contexto analítico***

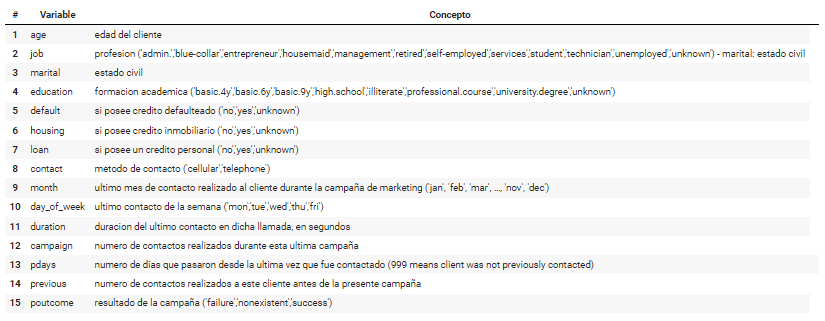
Se nos permitió el acceso a un .cvs donde emplearemos un análisis descriptivo para poder entender de mejor forma los datos recopilados y su relación con la variable a investigar "y" (Si el depósito a plazo fue realizado o no).

Dataset: <https://www.kaggle.com/datasets/rashmiranu/banking-dataset-classification>).

## 2. ****Dataset y Análisis de Variables****

El análisis servirá para verificar si una persona será propensa a realizar un depósito a plazo. La variable **output** a predecir es **y** donde 1: se suscribió 0: no se suscribió.

Para ello se cuenta con las siguientes variables **input** :

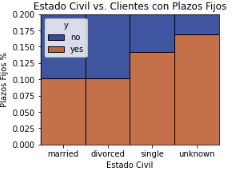


2.1 **Objetivos**

Las siguientes son las preguntas a responder a partir del conocimiento de los datos:

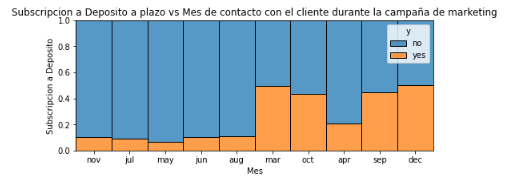
* Relación entre estado civil del cliente y Probabilidad de realizar depósitos a plazo
* Meses en que los clientes son más predispuestos a realizar depósitos a plazo
* Relación entre edad del cliente y duración del último contacto con el mismo con la Probabilidad que este realice un depósito a plazo

### 2.1.1 ****Relación entre estado civil del cliente y Probabilidad de realizar depósitos a plazo****



Se observa que predominan los depositos a plazo en clientes Solteros con un 15%, mientras que las personas casadas muestran un 10%, un 33% menor. Igualmente cabe resaltar que se desconoce el estado civil del 17% de las muestras.

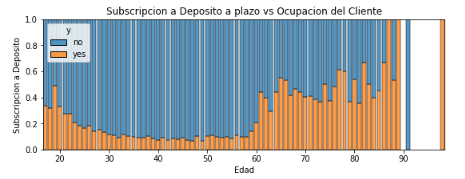
### 2.1.2 ****Meses en que los clientes son más predispuestos a realizar depósitos a plazo****



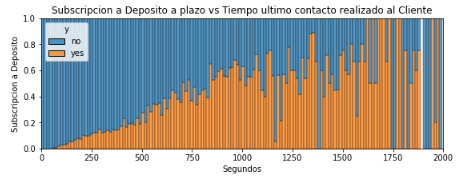
Podemos observar que los meses de Marzo, Octubre, Septiembre y Diciembre tuvieron un ratio sobresaliente de Subscripciones a Depositos a Plazo, un 200% mayor en promedio que el resto de meses.

### 2.1.3 ****Relación entre edad del cliente y duración del último contacto con el mismo con la Probabilidad que este efectué un depósito a plazo****

De la primera grafica podemos observar que los clientes con edad menor a 25 años y mayor a 60 años son más proclives a realizar depósitos a plazo.



De la segunda grafica se puede apreciar que las llamadas a clientes con duración menor a 100 segundos no concretaron con la realización de depósitos. Por otro lado, aquellas con una duración entre 760 segundos (12min) o más tuvieron más de un 50% de concretar en depósitos. Parece haber una correlación directa entre la ratio de suscripción y la duración de la llamada con el cliente.

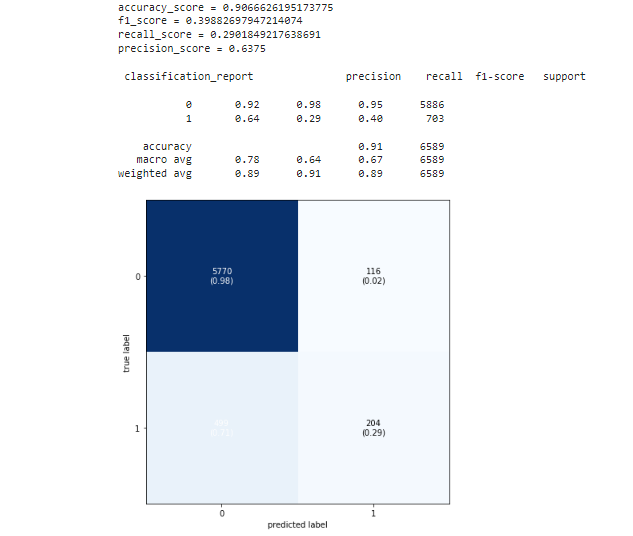


## 3. ****Desarrollo del Modelo****

A continuación, realizamos 3 métodos de machine learning (**Regresión Logística, Decision Tree y Random Forest**) y evaluamos su habilidad para predecir la predisposición de clientes a realizar depósitos a plazo en relación de las variables previamente analizadas.

### 3.1 ****Modelo por Regresión Logística****

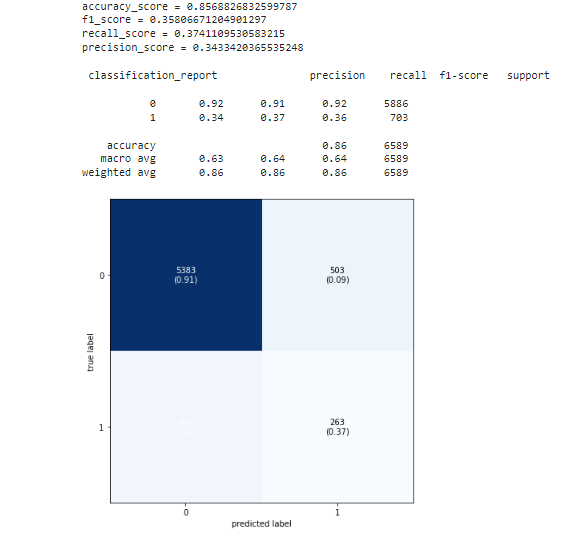
Utilizando los datos de duración de llamada, edad del cliente, el mes de contacto y estado civil del cliente realizamos un nuevo algoritmo de predicción, obteniendo un score de 91% utilizando un **Regresión Logística**.



De la matriz de confusión podemos observar la precisión del este preliminar modelo, observando los valores de falso y verdaderos negativos/positivos correspondientemente.

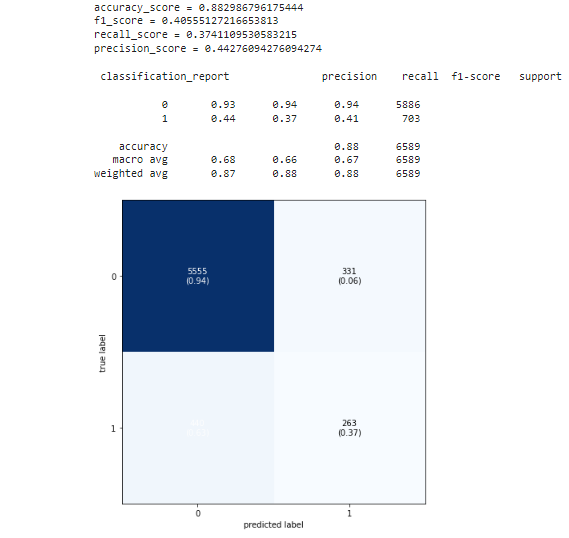
### 3.2 ****Modelo por Decision Tree****

Ahora utilizamos el método **Decision Tree**, obteniendo un score de 86%, inferior en 0.04 puntos porcentuales.



### 3.3 ****Modelo por Random Forest****

Ahora utilizamos el método **Random Forest**, obteniendo un score de 88%, inferior en 0.02 puntos porcentuales.



4.**Conclusiones**

Del análisis de los datos pudimos reconocer que las variables **Edad, Mes, Estado Civil y Tiempo de última llamada con el Cliente** juegan un factor determinante en la probabilidad de que los clientes se subscriban o no a un depósito a plazo. Siendo las siguientes las condiciones óptimas:

* Edad:18-25 y 60-100
* Mes: Marzo, Septiembre, Octubre y Diciembre
* Estado Civil: Soltero
* Tiempo de última llamada con el Cliente: 12min

Los modelos implementados lograron predecir el comportamiento observado en los datos con una precisión mayor al 86%, siendo el modelo de **Regresión Logística** aquel con mayor precisión, de un **91%**.

Proponemos utilizar el algoritmo de regresión logística en la base de datos de bancos para seleccionar a los prospectos a los cuales se le contactará mediante una campaña telefónica exclusiva, eficientizando el tiempo y el esfuerzo del call center.