Bank Marketing

Predicting subscription depending on background

Oscar Quesada Ávalos

UNIVERSIDAD CENFOTEC

Contents

[Abstract 2](#_Toc37021806)

[1. Limpieza de datos 2](#_Toc37021807)

[2. EDA 2](#_Toc37021808)

[3. RFE 4](#_Toc37021809)

[4. Matriz de Correlacion 5](#_Toc37021810)

[5. Modelos de Clasificacion 6](#_Toc37021811)

[6. Pregunta 7](#_Toc37021812)

# Abstract

El objetivo de la clasificación es predecir si el cliente suscribirá un depósito a plazo (variable y).

## Limpieza de datos

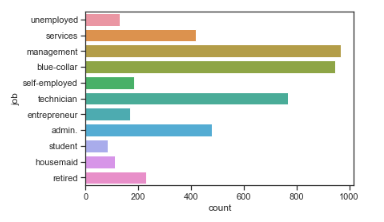
El data set posee 16 atributos y 4521 registros. Hacemos una revisión de nuestro data set para verificar si encontramos valores nulos o vacíos. Se revela que no se encuentra ese tipo de información. Se eliminan varias columnas contact-default-poutcome-balance, también se eliminan varias filas ya que tienen valores desconocidos es decir no clasificables.

## EDA

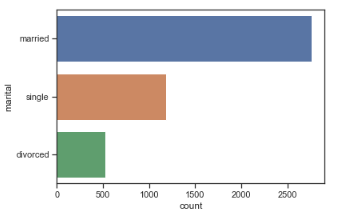
El dueño de la data set es Paulo Cortez (Univ. Minho) y Sergio Moro (ISCTE-IUL) 2012. Los datos están relacionados con campañas de marketing directo de una institución bancaria portuguesa. Las campañas de marketing se basaron en llamadas telefónicas. A menudo, se requería más de un contacto con el mismo cliente para acceder si el producto (depósito bancario a plazo) estaría (o no) suscrito.

Atributos

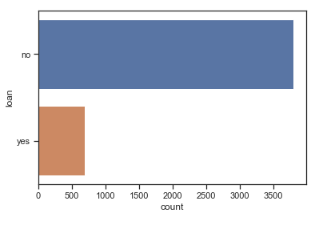
* age
* job
* marital
* education
* default
* balance
* housing
* loan
* contact
* day
* month
* duration
* campaign
* pdays
* previous
* poutcome
* y



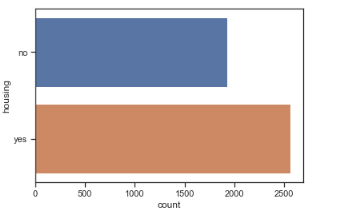
En el grafico anterior podemos verificar que las personas con el puesto de manager son los mayormente que poseen el producto.



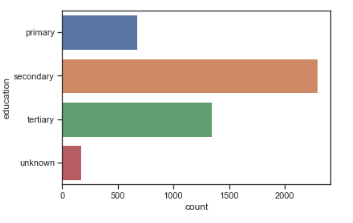
Vamos a verificar por estado marital quien posee el producto y se interpreta que con una predominancia son los casados.



Vamos a verificar por tener un crédito posee el producto y se visualiza que tenemos mayores clientes sin créditos que optan por el producto.

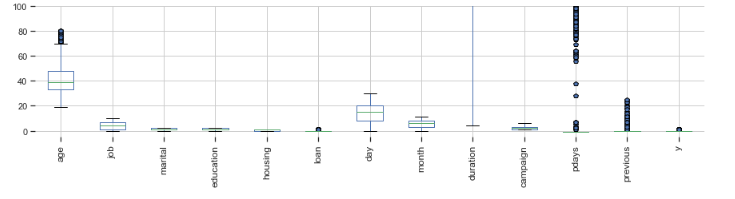


En este grafico Podemos mirar que en una gran mayoría las personas que no alquilan son los que generalmente tienen el producto. Es importante también visualizar la diferencia que hay entre ambos la cual es muy poca por eso se descartaría esta característica.



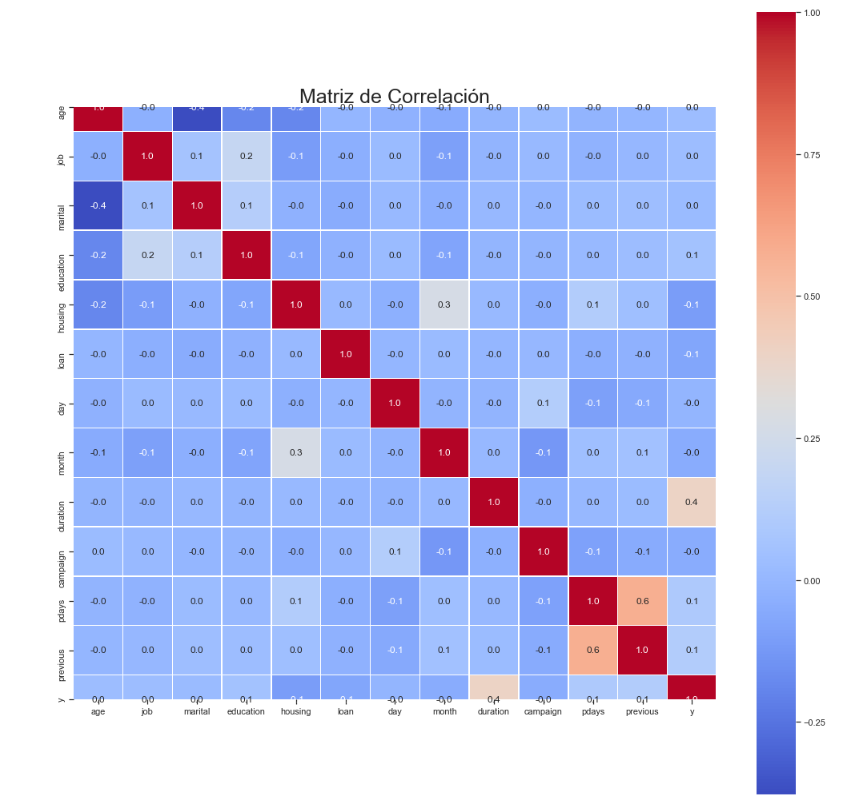
## RFE

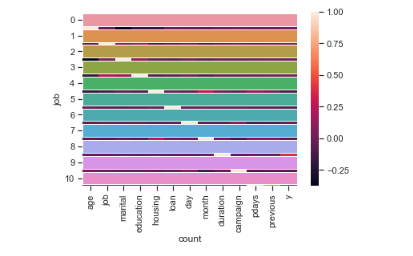
Podemos visualizar nuestro grafico con los atributos seleccionados para nuestra predicción.



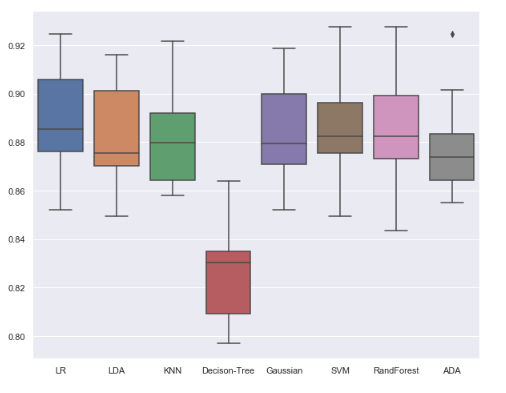
## Matriz de Correlacion

En dicha matriz Podemos visualizar una correlación que se encuentra pdays y previous para ambos ejes.





## Modelos de Clasificacion



Modelos Seleccionados:

**Logistic Regression** -- LR: 0.8894961240310076 (**Accuracy**)

**LDA**: 0.8825353892821031 (**Accuracy**)

**KNN**: 0.881370913380519 (**Accuracy**)

**Decison-Tree:** 0.8256917762049207 (**Accuracy**)

**Gaussian**: 0.8834049544994944 (**Accuracy**)

**SVM**: 0.8868849005729693 (**Accuracy**)

**RandForest**: 0.8863060330299966 (**Accuracy**)

**ADA**: 0.8784706774519717 (**Accuracy**)

Modelo seleccionado 🡪 **Logistic Regression con mayor precisión.**

## Pregunta

Predecir si el cliente suscribirá un depósito a plazo (variable y)?

El cliente actual que poseemos en nuestros datos no va subscribir un deposito a plazo, por tanto debemos analizar los diferentes perfiles para tener identificadas las caracterizaciones de un cliente si va subscribir el producto.