



Travel Insurance

Autor: Leandro Hornos

Nuestro cliente



AGENCIA
DE VIAJES



INDIA



2000
CLIENTES

El Proyecto

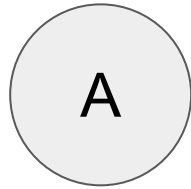


En 2019 la empresa ofreció un **seguro de viajes** a sus clientes

La empresa quiere conocer el **perfil de los clientes** que se interesan por seguros de viaje

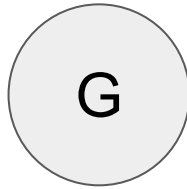
Se desea construir un **modelo** de machine learning capaz de identificar aquellos clientes con mayor potencial de **adquirir el seguro**.

Datos de los clientes



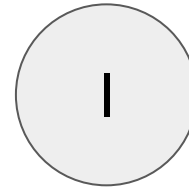
AÑOS

Edad del cliente



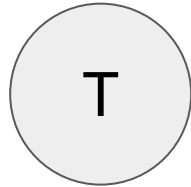
GRADUACIÓN

Si el cliente es graduado universitario o no



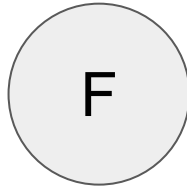
INGRESO ANUAL

Cantidad de rupias anuales que percibe el cliente



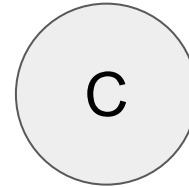
TIPO DE EMPLEO

Público o privado



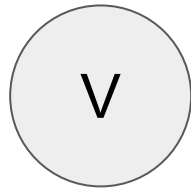
FAMILIARES

Cuántas personas componen el grupo familiar del cliente



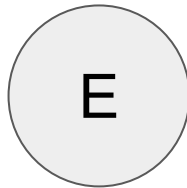
ENFERMEDAD CRÓNICA

Presenta o no una condición crónica



VIAJERO FRECUENTE

Más de 4 reservas en los últimos 2 años



VIAJÓ AL EXTERIOR

Ya sea con esta empresa u otra.



ASEGURADO

Si adquirió un seguro o no en 2019.

Características generales de los clientes

Los clientes de la agencia son **personas jóvenes** de entre 25 y 35 años con una media de 29 años, y que encabezan **familias** que se componen de unos 4 a 5 miembros.

En general son personas saludables, y en su mayoría no poseen mucha experiencia en viajes

85%

Graduados

28%

Enfermedades
Crónicas

19%

Viajó al
Exterior

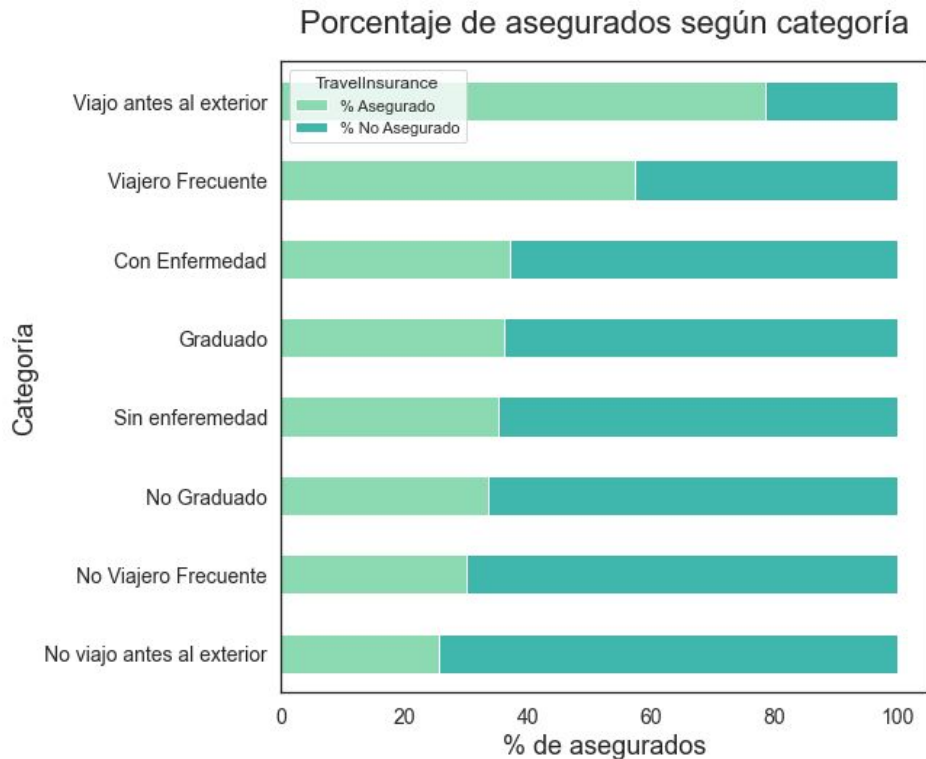
21%

Viajero
Frecuente

En 2019, el 36% adquirió un seguro de viaje



Características vs. Aseguramiento



Dividimos a los clientes entre aquellos que cuentan o no cuentan con determinada característica y analizamos el porcentaje de asegurados en cada categoría

La mayor proporción de asegurados está en quienes han viajado antes al exterior.

Contrariamente, entre quienes no lo han hecho está la menor proporción de asegurados.

Modelo 1: Árbol de decisiones

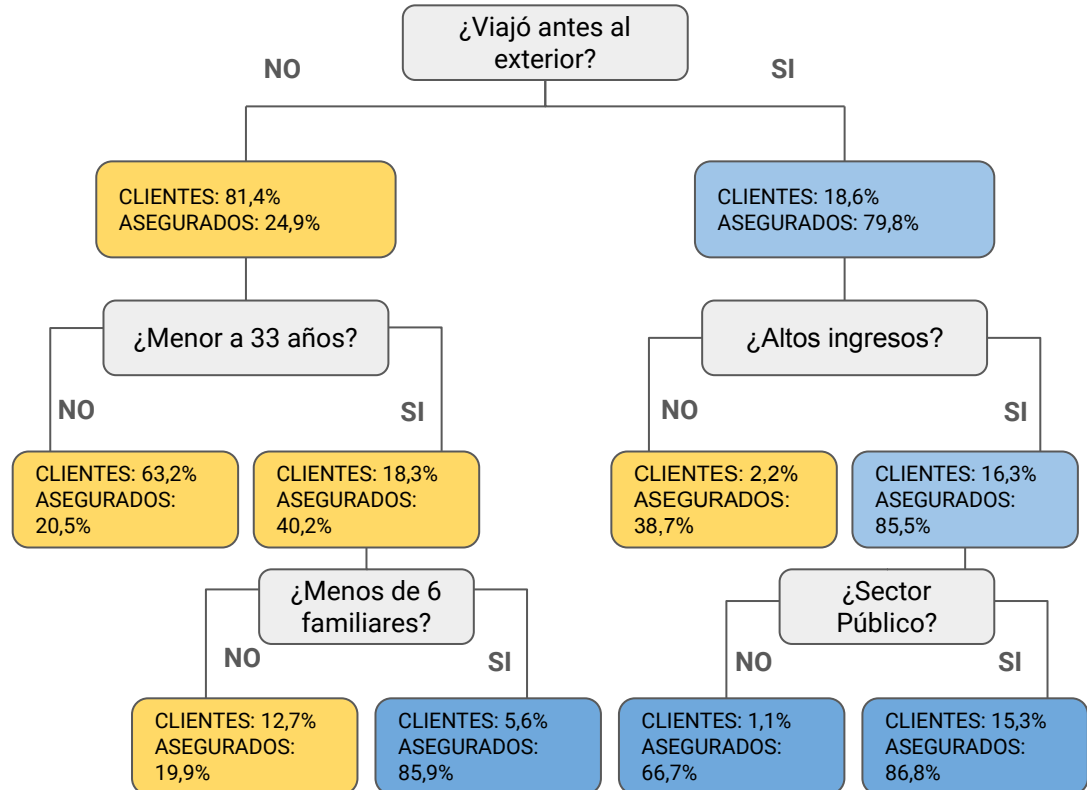
Como primera aproximación al modelado se optó por un árbol de decisiones.

Este modelo nos permite visualizar claramente las características que definen a los potenciales compradores del seguro

CLIENTES: % del total de clientes
ASEGURADOS: % asegurado dentro del subgrupo

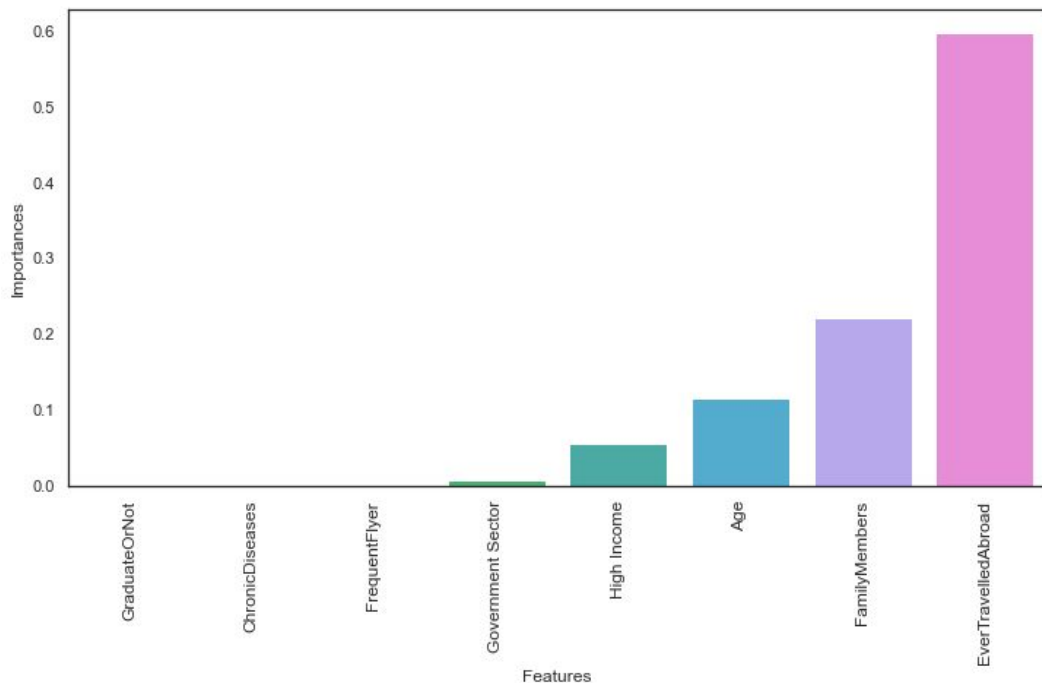
AZUL: MAYORÍA DE ASEGURADOS

NARANJA: MAYORÍA DE NO ASEGURADOS



Modelo 1: Árbol de decisiones

Decision Tree: Importancia de cada feature



De acuerdo con lo visto en el análisis exploratorio el modelo le da principal importancia a que el cliente haya viajado al exterior alguna vez, así como a la edad y la composición familiar

Métricas del modelo

Precision: 88%

Recall: 53%

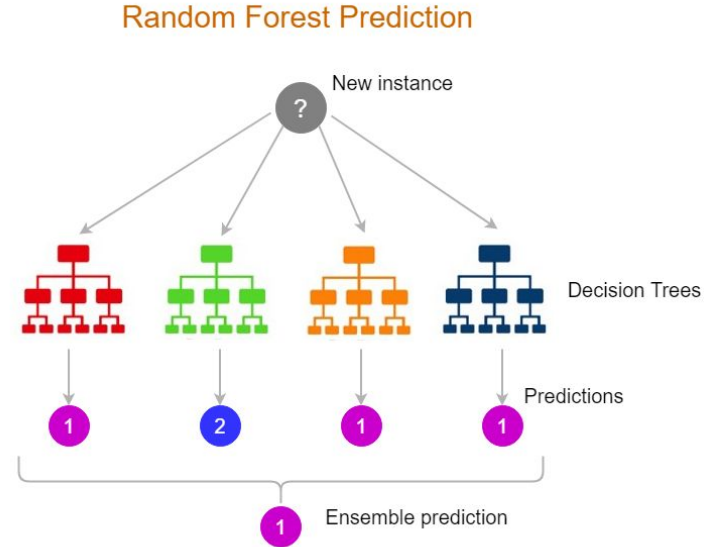
El modelo genera muy pocos falsos positivos, pero genera muchos falsos negativos.

Es bueno para minimizar pérdidas pero no para maximizar ventas

Modelos alternativos

Con el objetivo de obtener un modelo que genere una menor tasa de falsos negativos, se prueban 3 modelos alternativos:

- **Random Forest**
- **Adaboost**
- **XGboost**

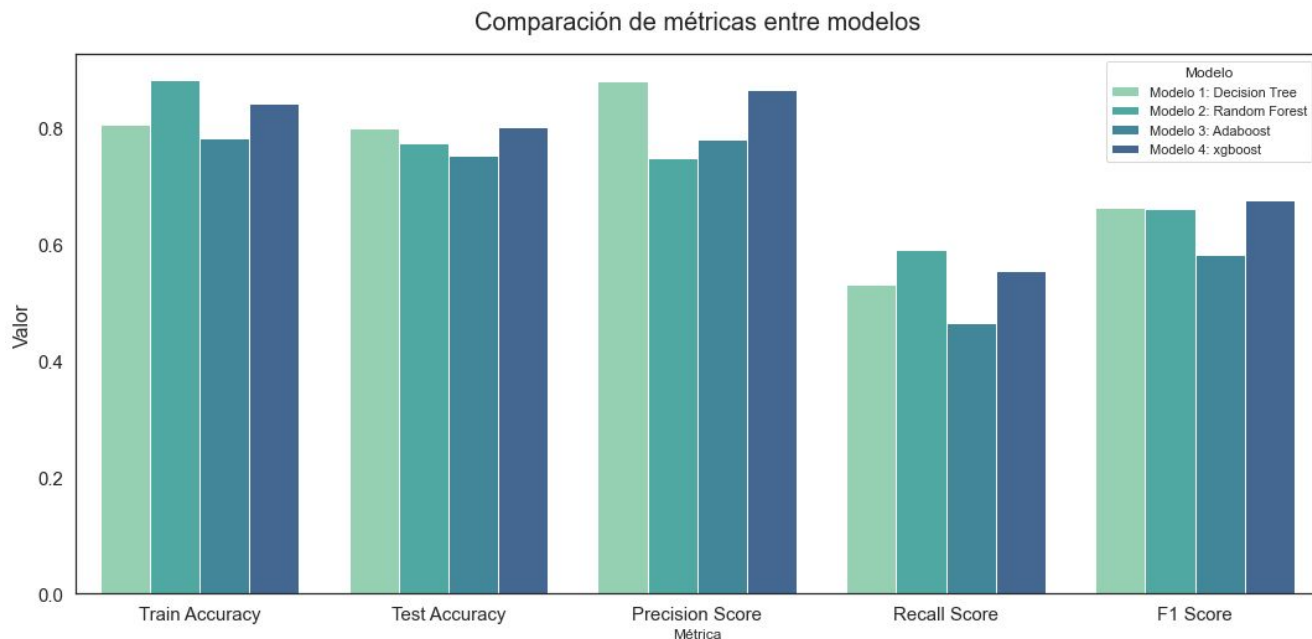


XGBoost

AdaBoost

Modelos alternativos

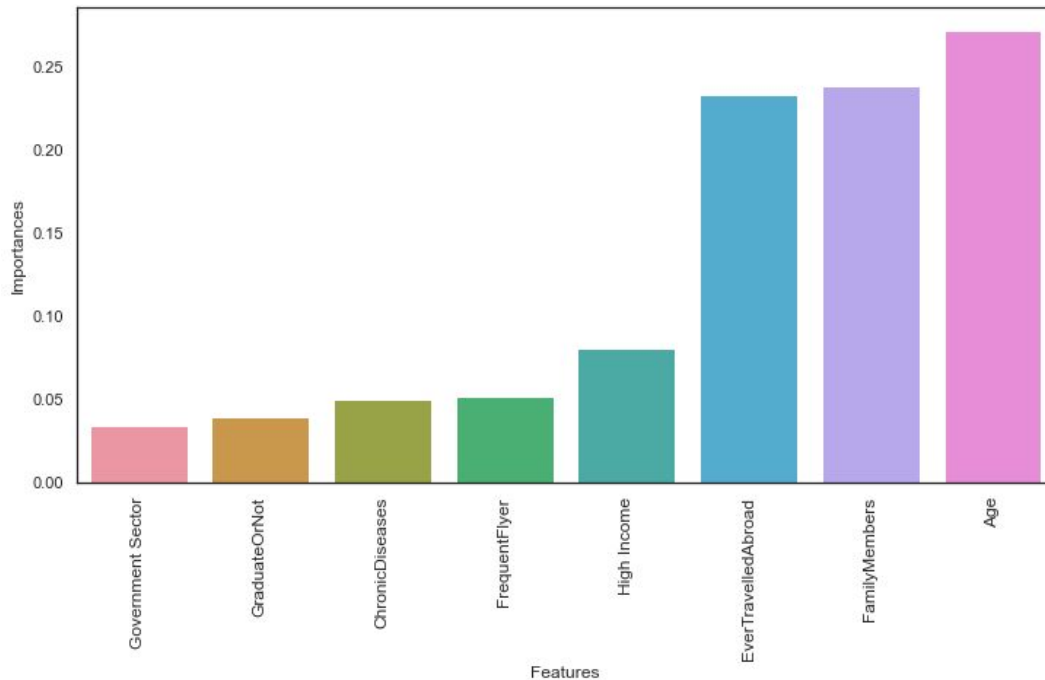
Se elige el modelo de Random Forest ya que muestra el mejor recall, y es una buena mejora respecto al árbol de decisiones con un modelo relativamente simple



Modelo 2: Random Forest

El modelo de random forest sigue dándole importancia al hecho de que el cliente haya viajado alguna vez al exterior, pero da más peso a la edad y la composición familiar que el árbol de decisiones sencillo

Random Forest: Importancia de cada feature



Conveniencia económica del modelo

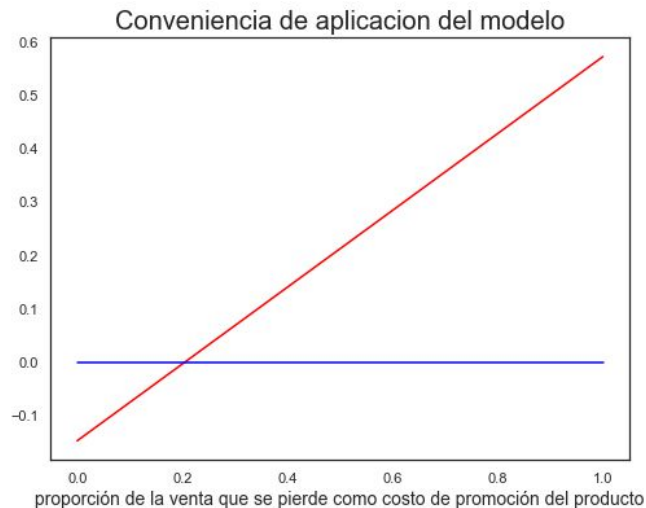
La ganancia obtenida aplicando el modelo es mayor a la de no aplicarlo si se cumple que

$$P(\text{Recall} - 1) + X(1 - P \frac{\text{Recall}}{\text{Precision}}) > 0$$

P es la proporción de personas que adquirieron el seguro

X es la fracción de la ganancia de la venta de un seguro que se pierde como costo de promoción

El modelo comienza a ser conveniente cuando la recta pasa por 0



Conviene ofrecer el seguro sólo a aquellos clientes identificados por el modelo cuando el costo de promoción del seguro supera el 20% de la ganancia por venta

Conclusiones

- Se estudiaron las características de los clientes que adquieren el seguro y se detectó que aquellas personas que **han viajado antes al exterior** son las que muestran mayor interés en seguros de viaje
- Se desarrolló un **modelo de machine learning** (Random Forest) capaz de clasificar a los clientes e identificar aquellos con alto potencial de adquirir el seguro
- El modelo desarrollado es mucho más eficiente **minimizando las pérdidas** que maximizando las ventas, ya que genera pocos falsos positivos pero pasa por alto una gran proporción de los potenciales compradores
- Se mostró que **el modelo es económicamente ventajoso cuando el costo de promoción del seguro es una fracción elevada de la ganancia obtenida por la venta**, y se estimó que cuando dicho costo supera el 20% es conveniente aplicar el modelo
- Respecto a un potencial interés frente a un seguro de cobertura por COVID-19, vemos que el dataset no cuenta con información relevante al respecto como para poder hacer alguna clase de predicción.

MUCHAS GRACIAS!