



**PORTO
SEGURO**

PORTO SEGURO'S

CLAIM PREDICTION



Presented by Daniel Vega

3,6% de los clientes reclaman

PREDECIR LOS CLIENTES QUE
RECLAMARÁN EL AÑO QUE VIENE



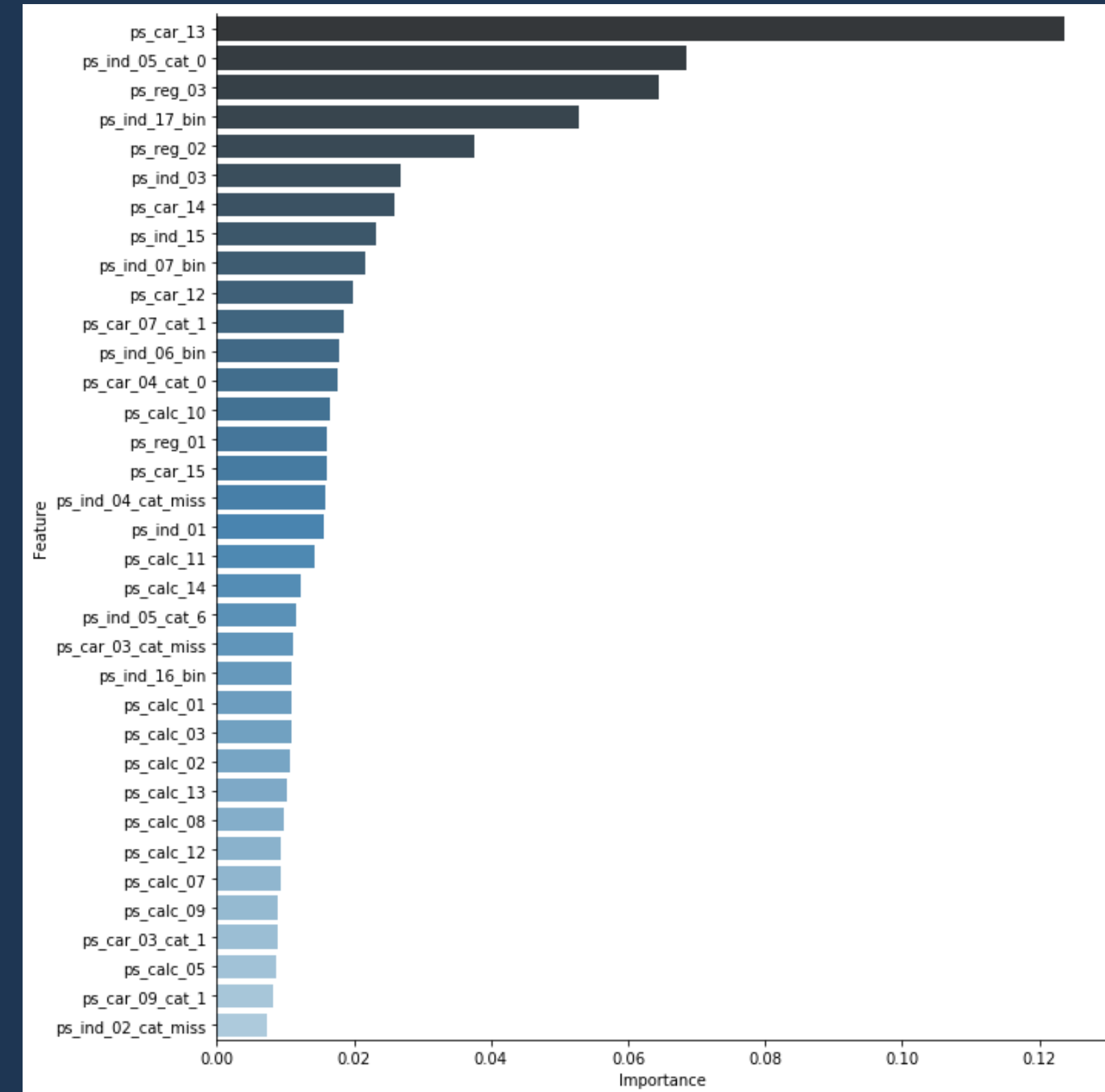
Con esta información se podrían adaptar las tarifas del seguro a cada tipo de cliente y hacer que la cobertura del seguro de automóviles sea más accesible para más conductores.

CONOCE A TUS CLIENTES Y AJÚSTATE A ELLOS!



DATOS

- Data set con casi 600.000 registros de clientes diferentes.
- Cada registro contiene información de 57 variables diferentes en 4 categorías: individuo, coche, regional y calculadas.
- Las variables están encriptadas.

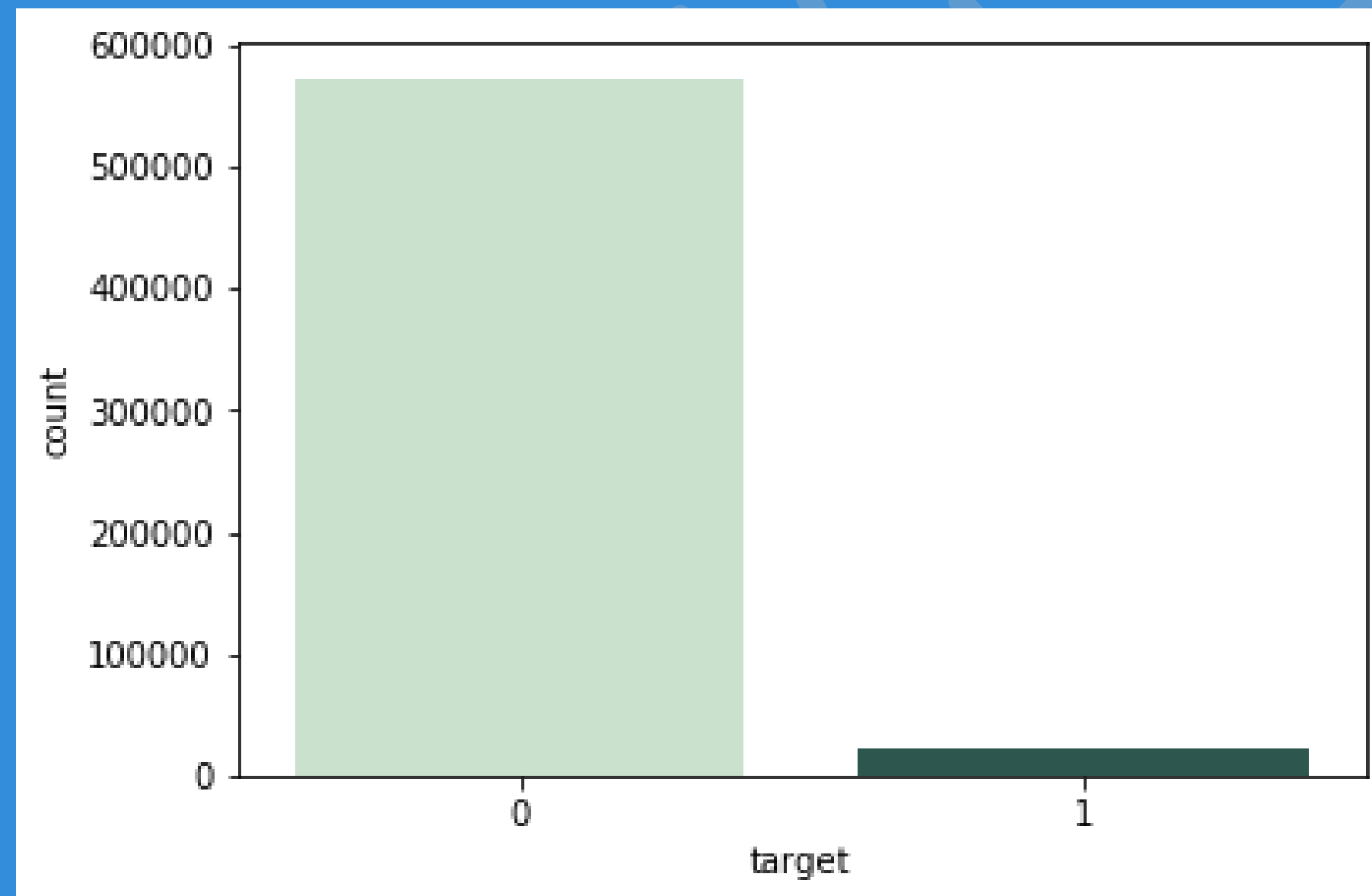


TARGET

DESBALANCEADO,
SOLO 3,6%
REGISTROS DE
RECLAMACIÓN

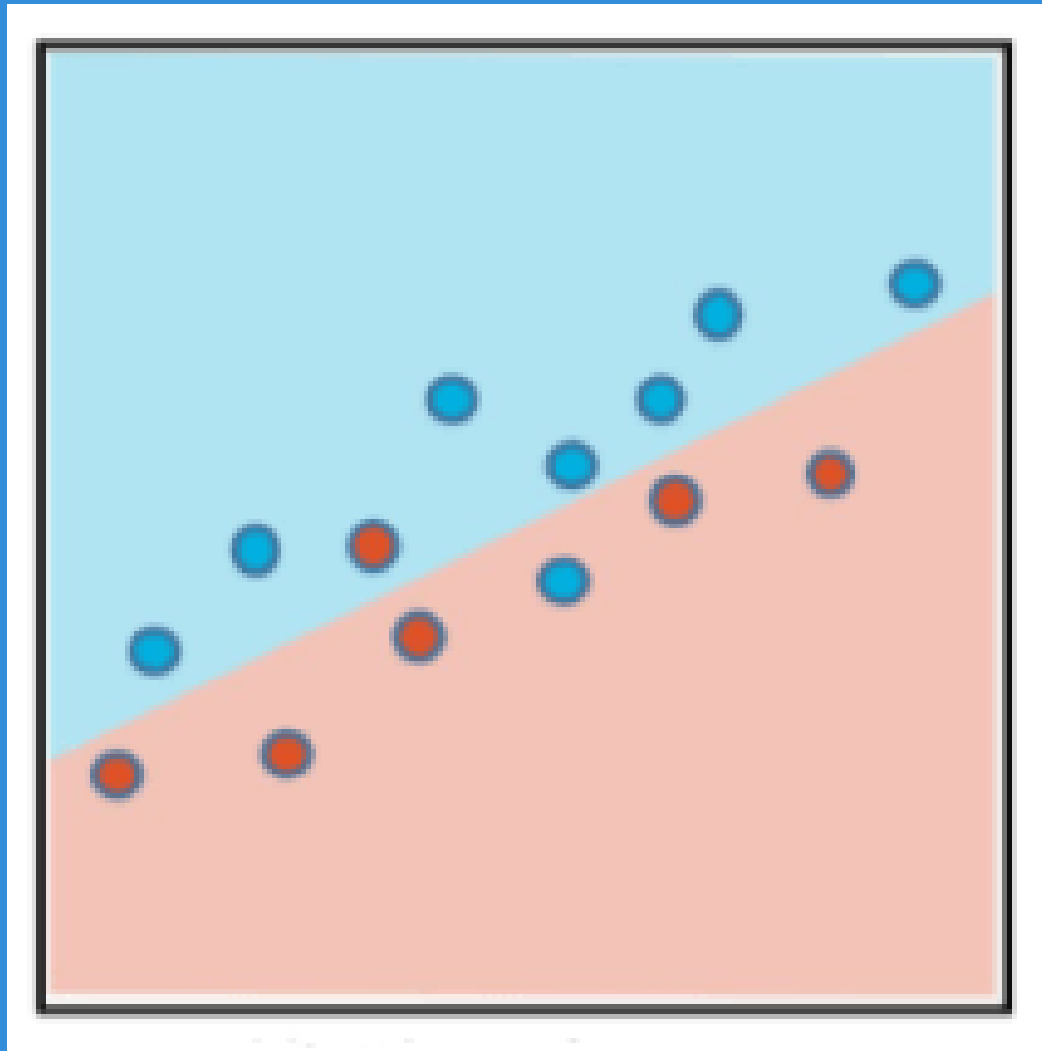
0 = no reclama

1 = reclama



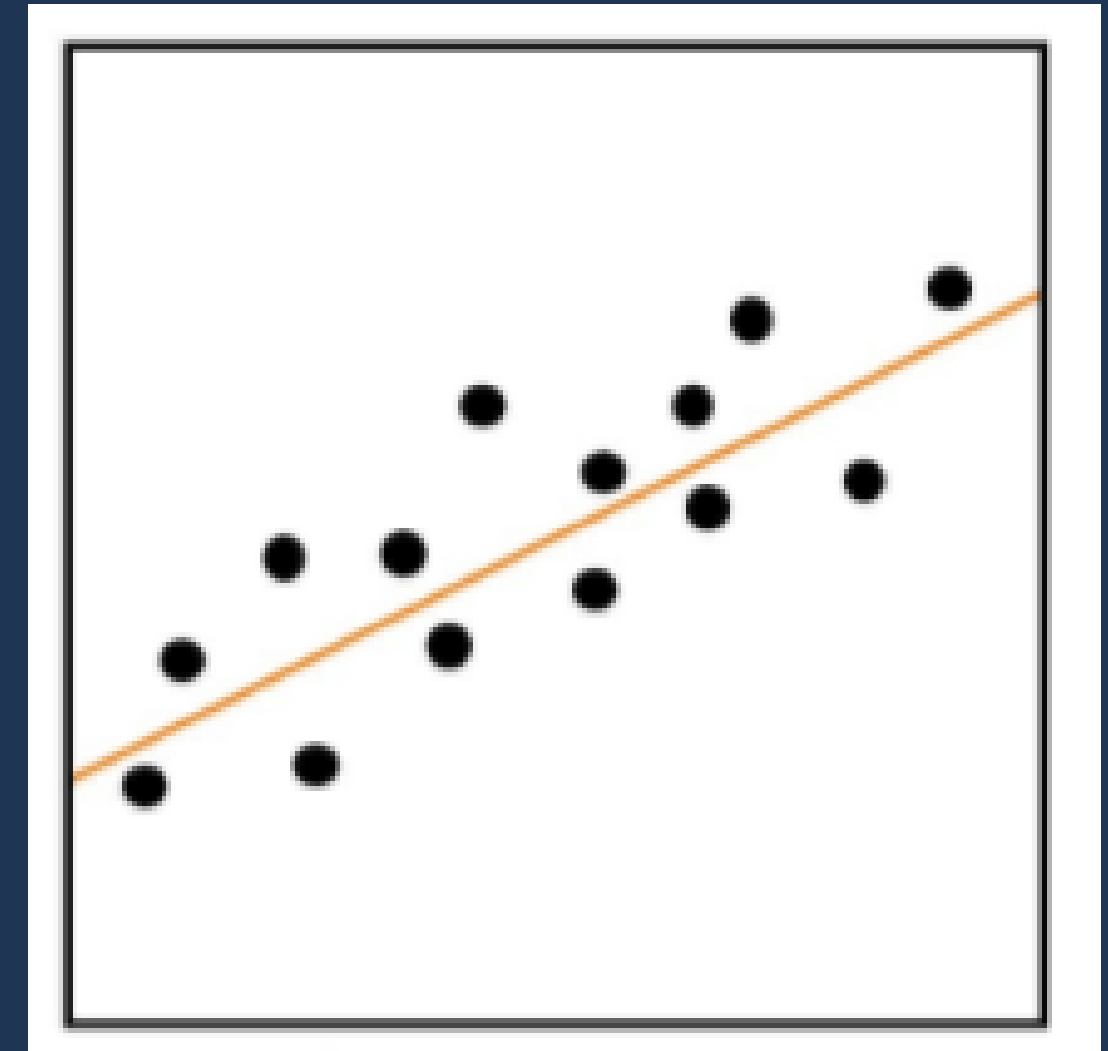
QUÉ ES UN MODELO?





CLASIFICACIÓN

El modelo nos devuelve las probabilidades de pertenecer a cierta categoría.

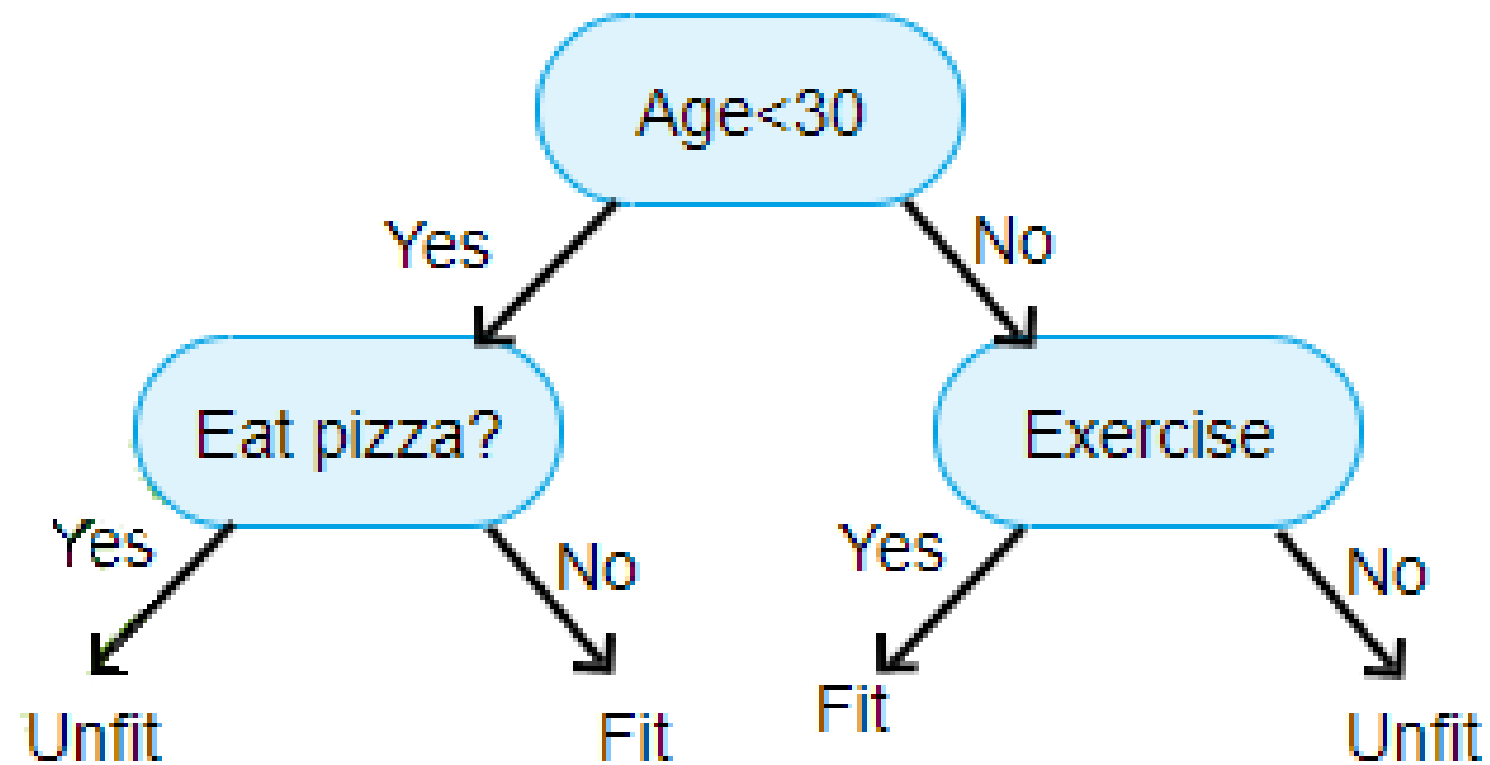


REGRESIÓN

El modelo nos devuelve las predicciones en forma de valor numérico.

Árbol de decisiones

- El árbol va seleccionando las variables óptimas y segmentando los registros en base a una condición.
- En las hojas caen los registros clasificados en base a una probabilidad.



MODELOS



ÁRBOL

limonero



RANDOM FOREST

la media de los limones de muchos
limoneros diferentes



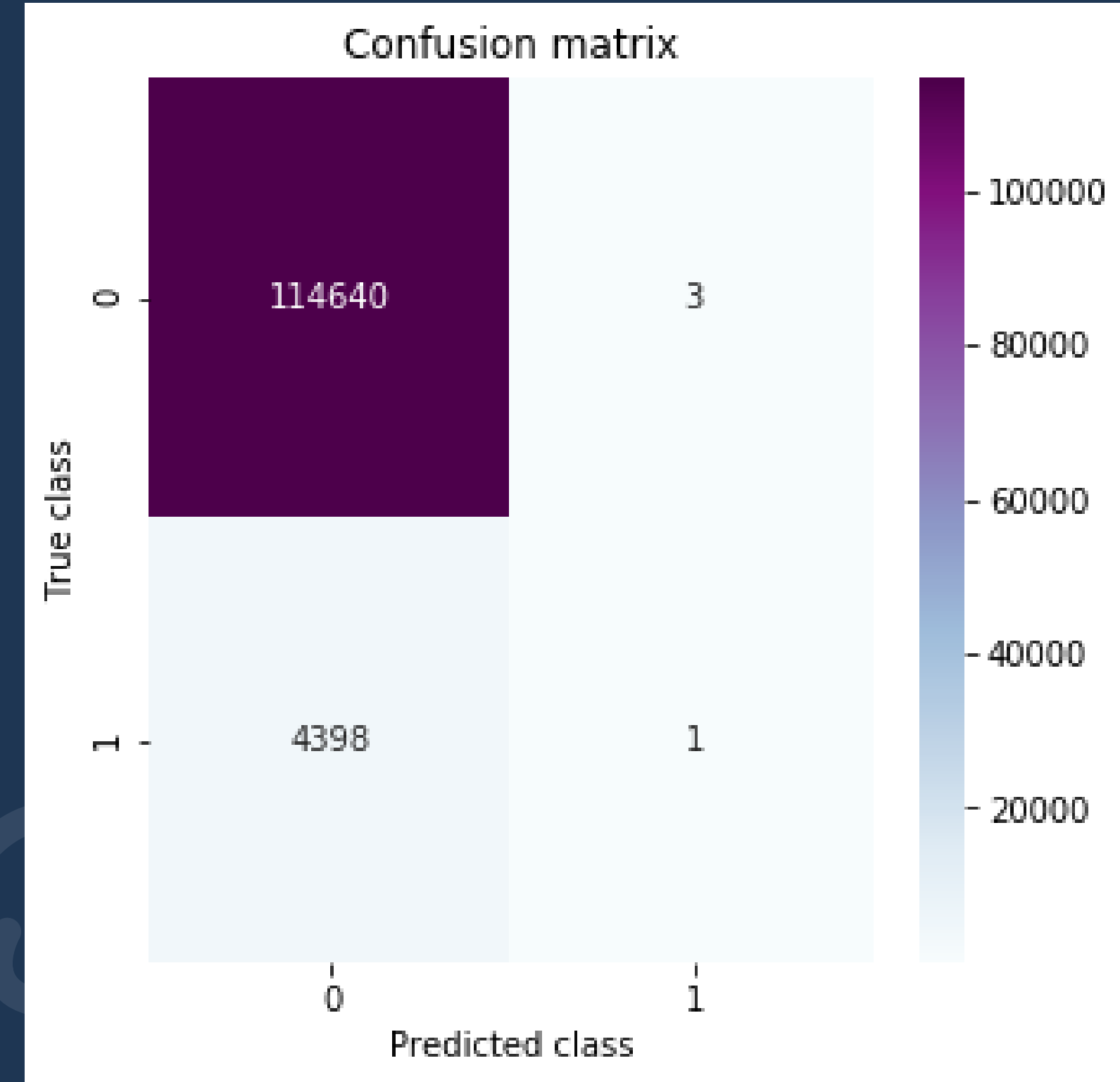
LIGHTGBM

limoneros listos, aprenden del error
de sus vecinos

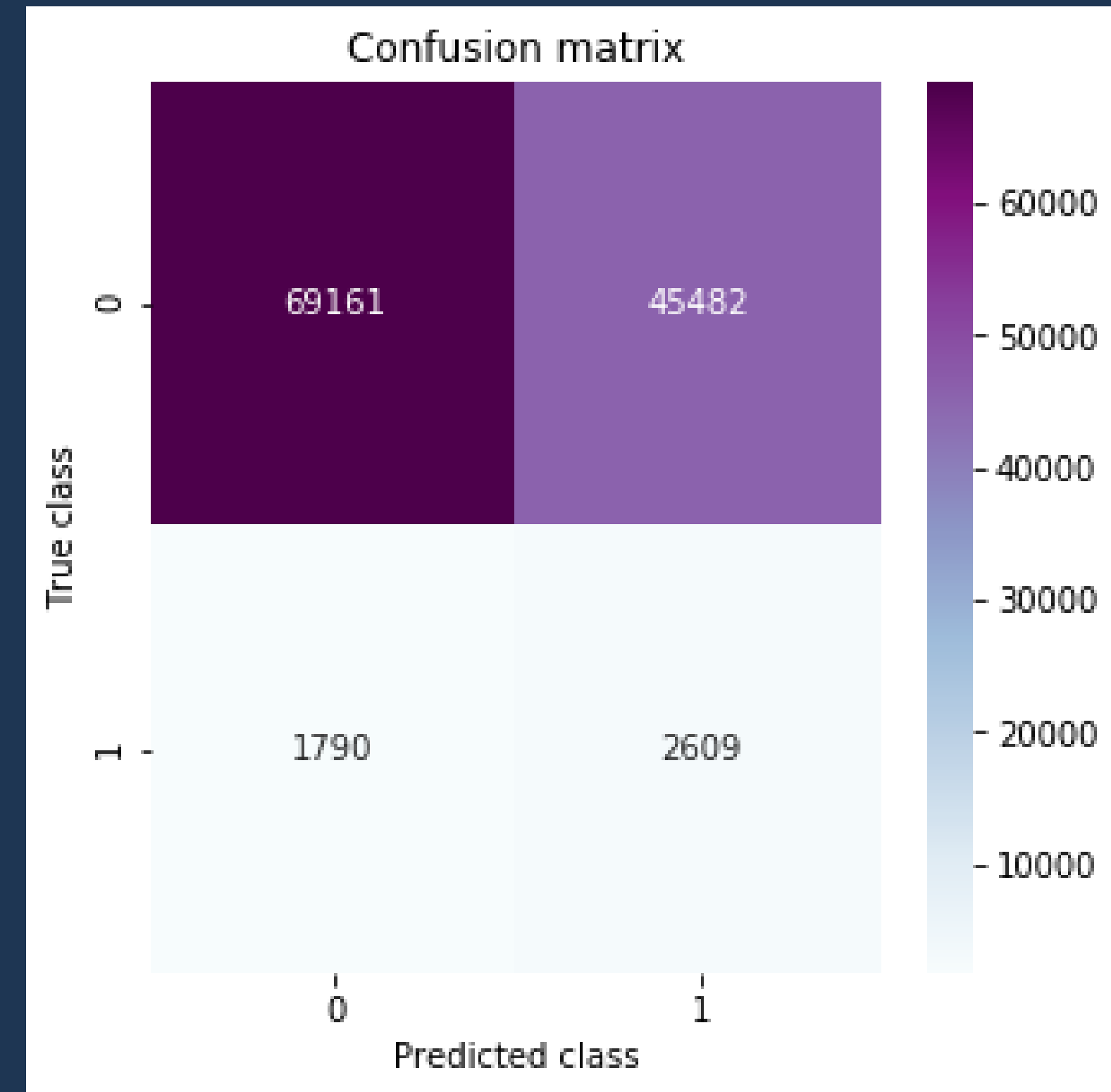
RESULTADOS

96,4% de precisión en la predicción.

- El modelo no identifica casi ningún cliente como evidente reclamador potencial.
- Podemos corregirlo ajustando el parámetro que se encarga de clasificar las predicciones, el Threshold.
- Nos basaremos en otras métricas que evalúan los aciertos para diferentes valores del Threshold, o punto de corte.

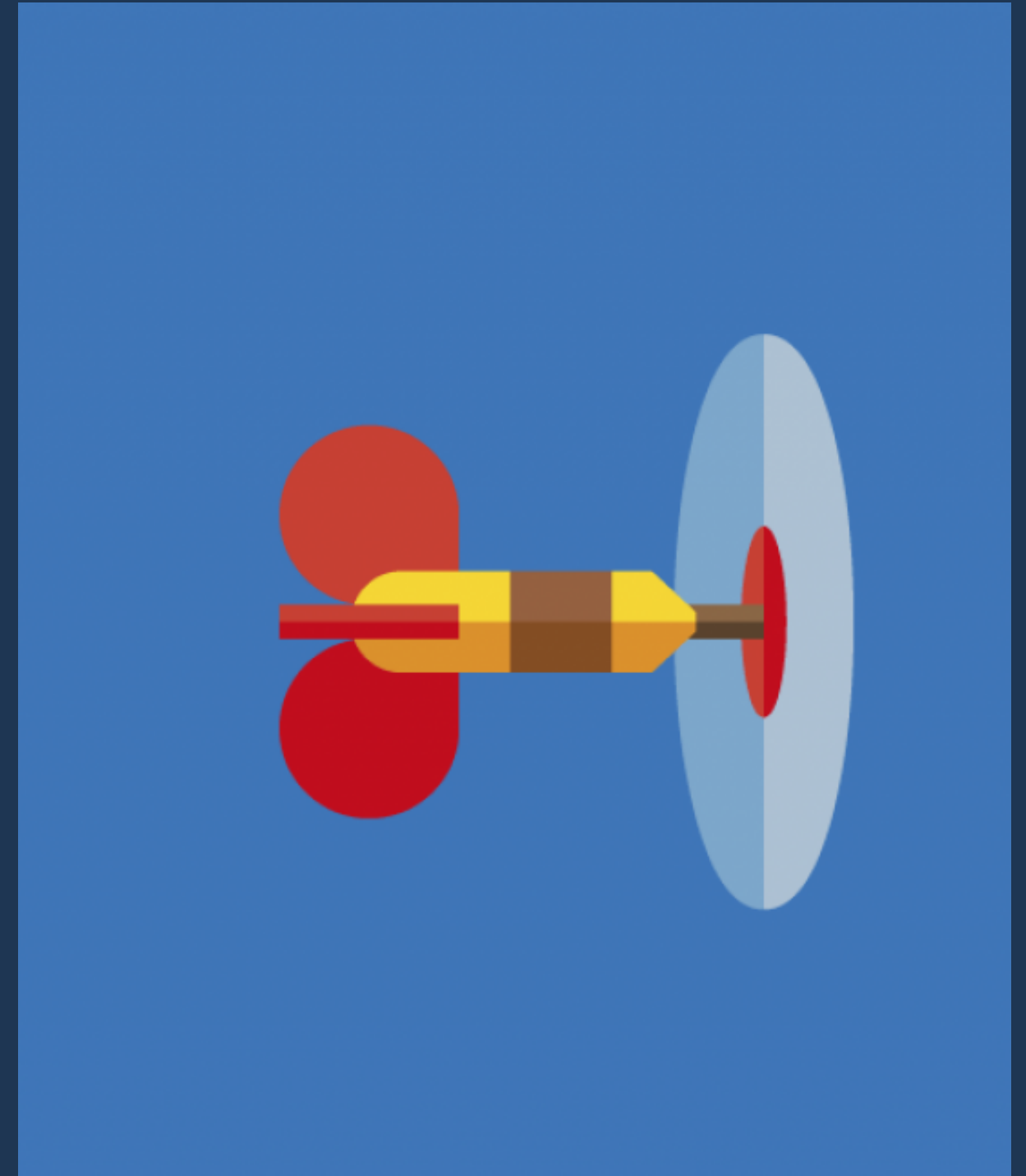


- Hemos predecido el comportamiento de 119042 individuos.
- Ajustando el Threshold a un valor de: 0.35 que maximiza el número de aciertos en los casos positivos y negativos obtenemos el siguiente resultado:



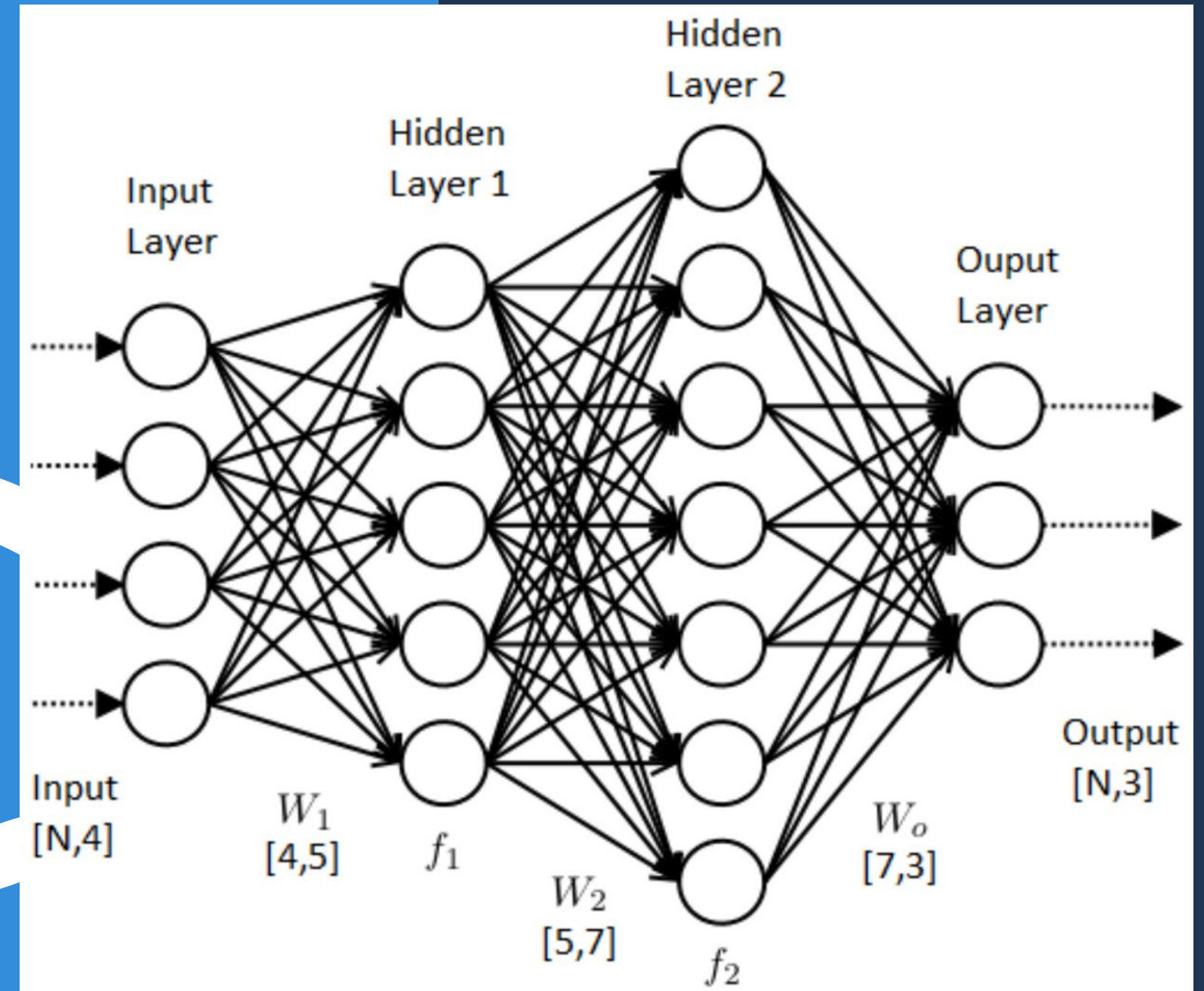
CONCLUSIONES

- No tenemos datos de calidad suficiente para identificar con precisión a los reclamadores.
- Conocer información acerca de las variables ayudaría mucho a la hora de tratar y crear nuevas variables para mejorar la predicción.



Extras:

- En este problema sería interesante ajustar otro tipo de modelo como una red neuronal.



GRACIAS

