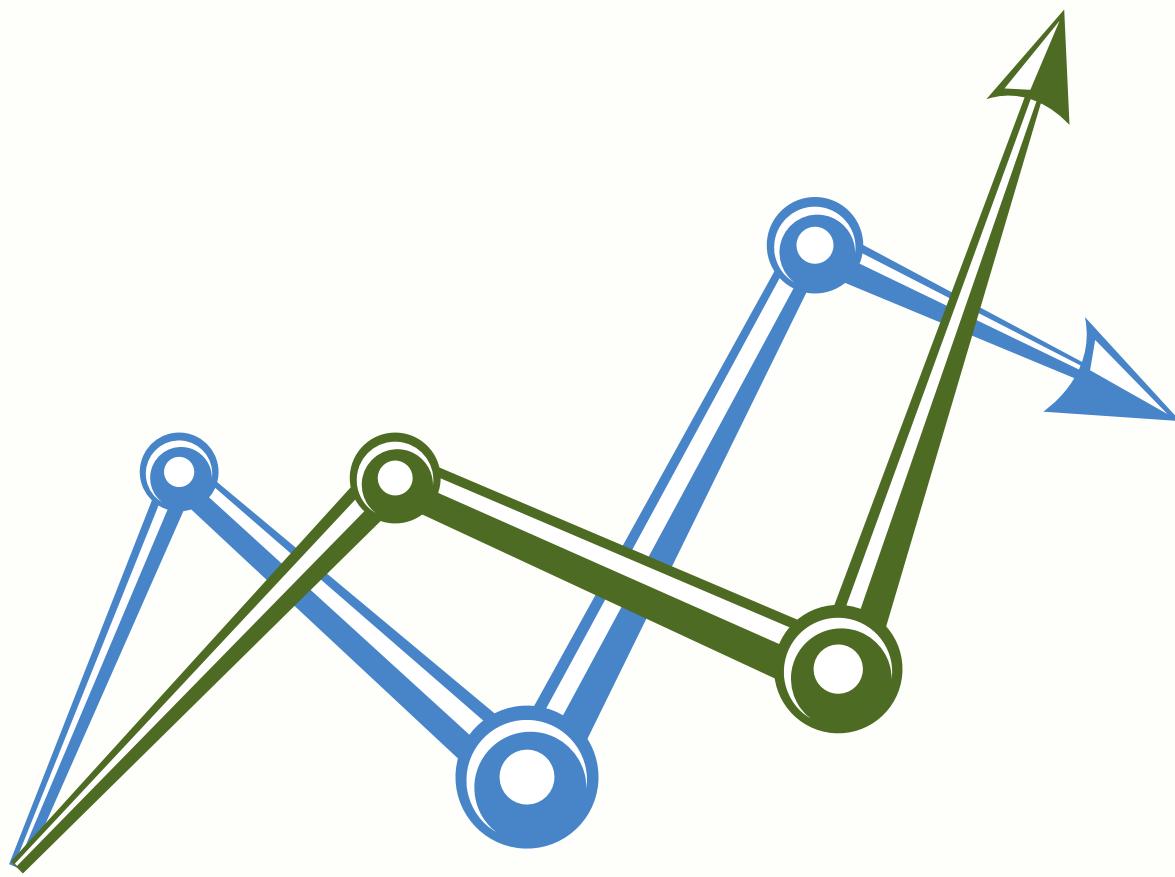


**RETO**

JIN WANG  
HUGO GUTIÉRREZ  
JIAYI WU  
PAULA AGORRETA

Propuesta de predicción de ventas

**Schneider**  
 **E l e c t r i c**



# EL PROBLEMA

Nuestra misión consiste en construir un **algoritmo** que prevea si se va a **ganar oportunidades de venta**, o si se van a **perder**, también aplicar **técnicas de explicabilidad** para analizar los datos extraídos

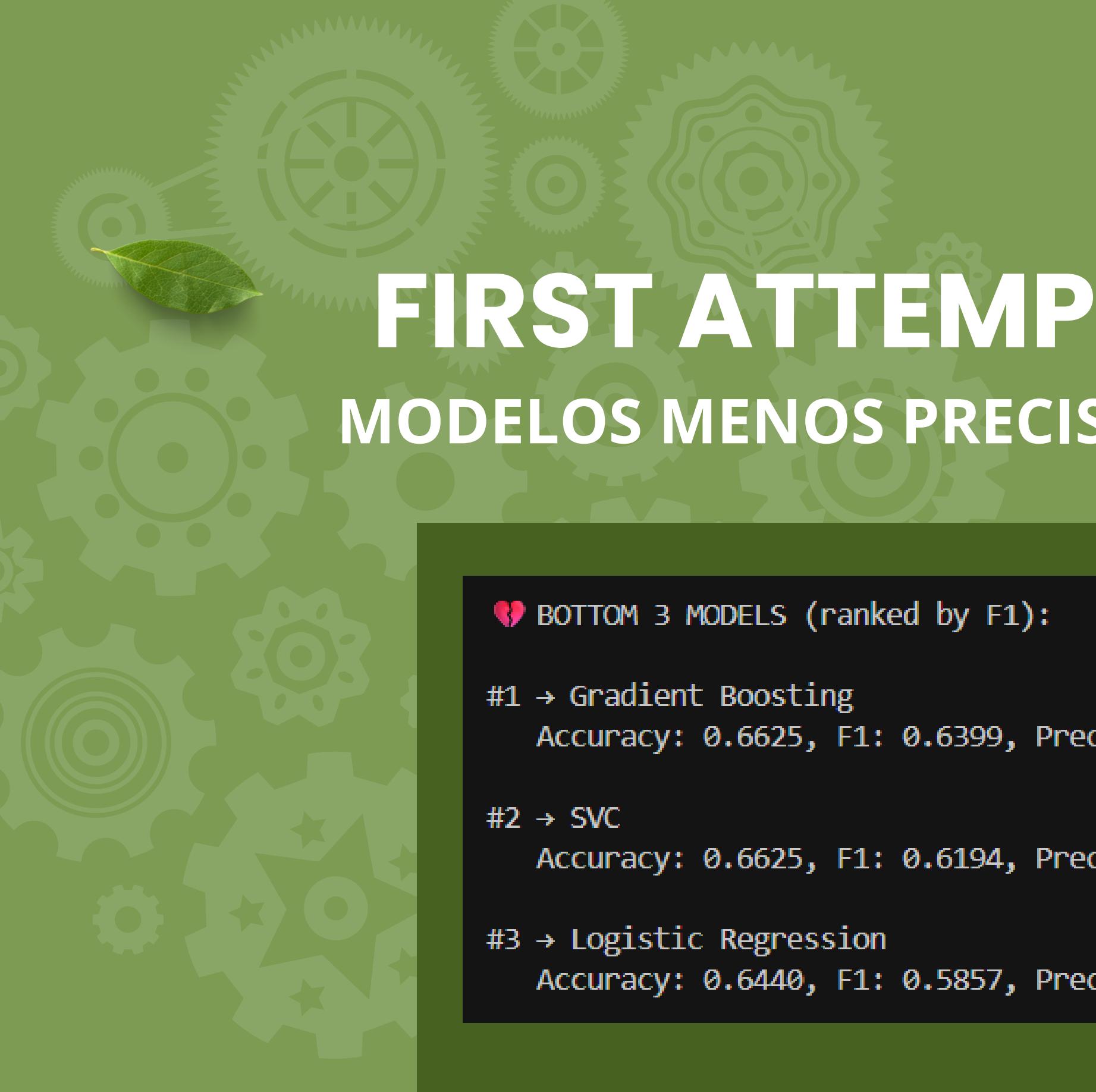
# FASE INICIAL

- ENCONTRAR LIBRERIAS
- INFORMARNOS DE LAS INSTRUCCIONES
- CONSULTA DEL LIBRO DE STATEMENTS



# Selección de modelos ML

Con este propósito,  
comparamos 10 modelos en  
base a **velocidad y precisión**



# FIRST ATTEMPT

## MODELOS MENOS PRECISO

# Selección de modelos ML

💔 BOTTOM 3 MODELS (ranked by F1):

- #1 → Gradient Boosting  
Accuracy: 0.6625, F1: 0.6399, Precision: 0.6321, Recall: 0.6479, Time: 2.537986s
- #2 → SVC  
Accuracy: 0.6625, F1: 0.6194, Precision: 0.6478, Recall: 0.5934, Time: 197.530233s
- #3 → Logistic Regression  
Accuracy: 0.6440, F1: 0.5857, Precision: 0.6347, Recall: 0.5438, Time: 0.022063s

en  
ión

# VELOCIDAD

#1 → Naive Bayes

Time: 0.01s, Accuracy: 0.5166,

#2 → Logistic Regression

Time: 0.02s, Accuracy: 0.6440,

#3 → KNN

Time: 0.08s, Accuracy: 0.7765,

#4 → XGBoost

Time: 0.22s, Accuracy: 0.7705,

# Selección de modelos ML

Con este propósito,  
comparamos 10 modelos en  
base a **velocidad y precisión**

# PRECISIÓN RANKING POR F1

🏆 TOP 4 MODELS (ranked by F1):

#1 → Random Forest  
Accuracy: 0.8398, F1: 0.8266,

#2 → KNN  
Accuracy: 0.7765, F1: 0.7563,

#3 → XGBoost  
Accuracy: 0.7705, F1: 0.7511,

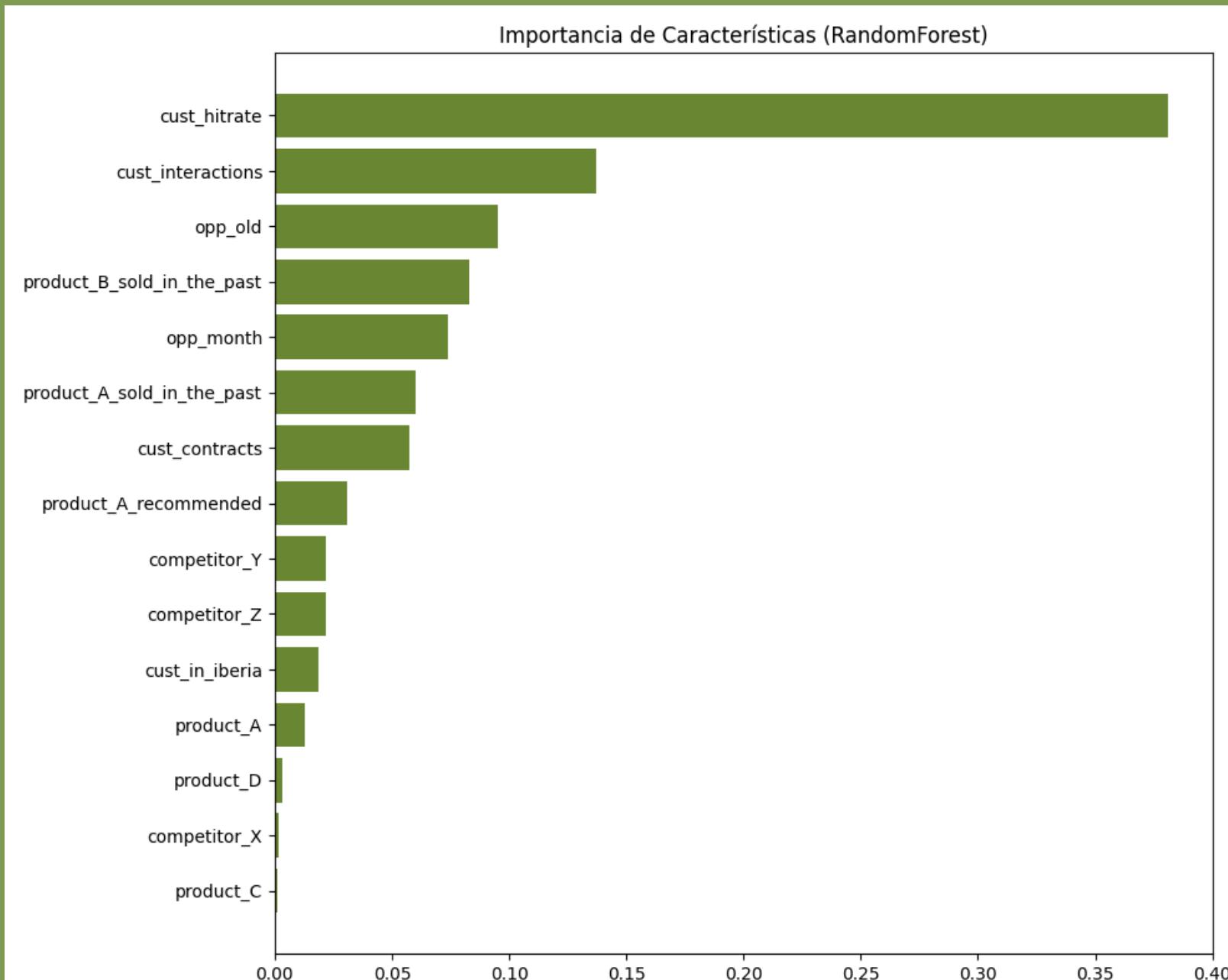
# Selección de modelos ML

Con este propósito,  
comparamos 10 modelos en  
base a **velocidad y precisión**



# DATOS

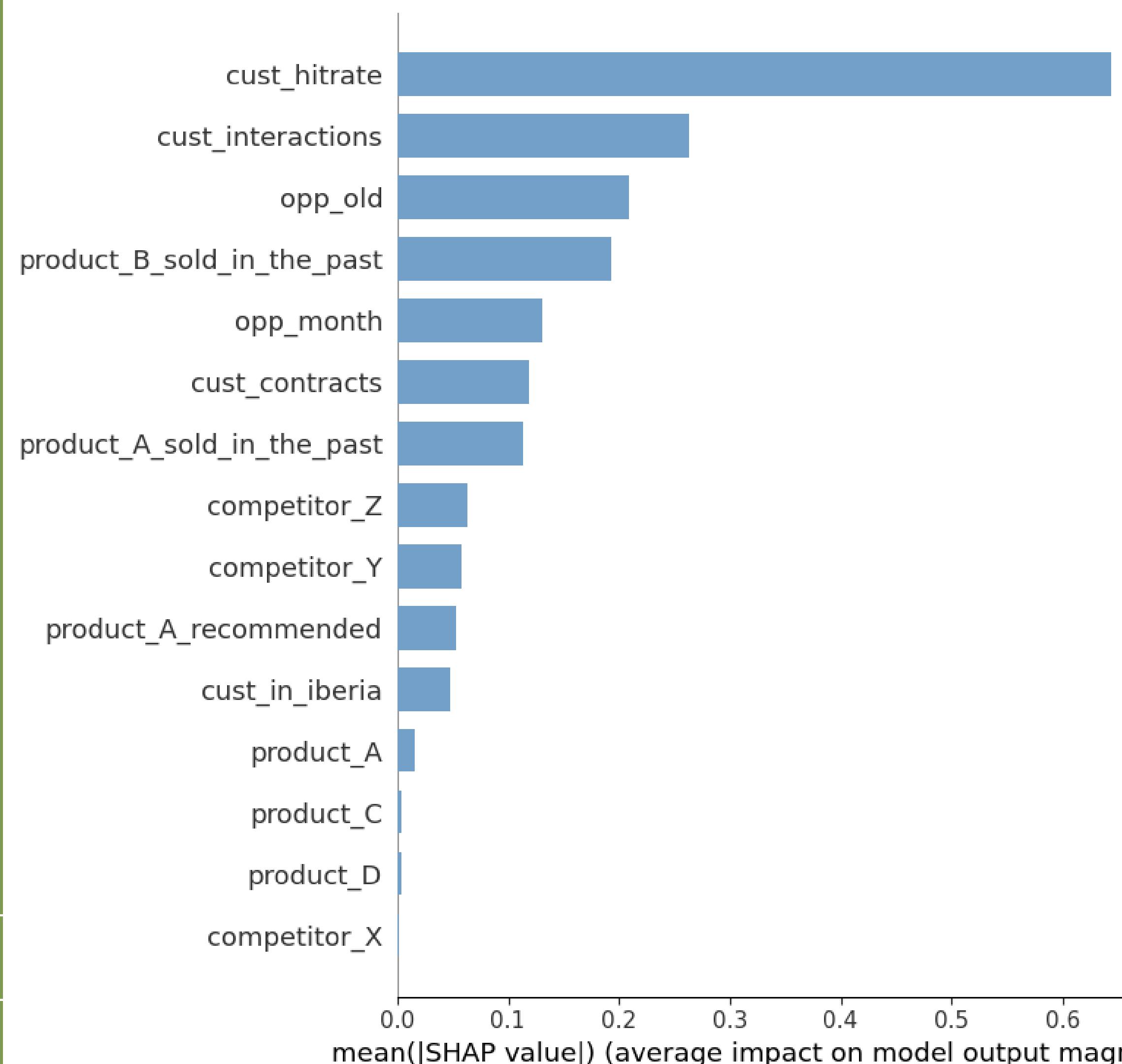
# IMPORTANCIA DE ATRIBUTOS



# XGBOOST



FLOP 3

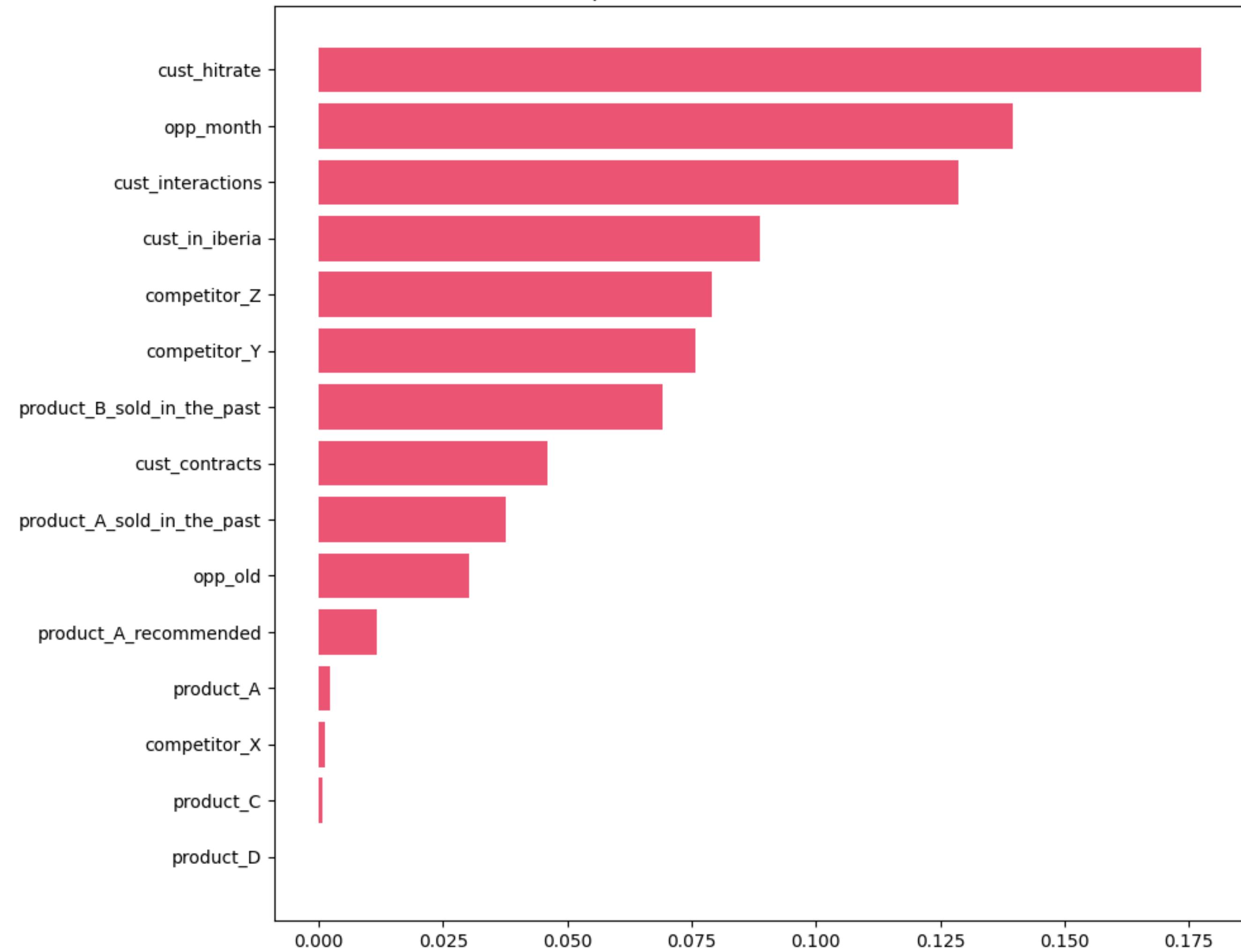


# KNN



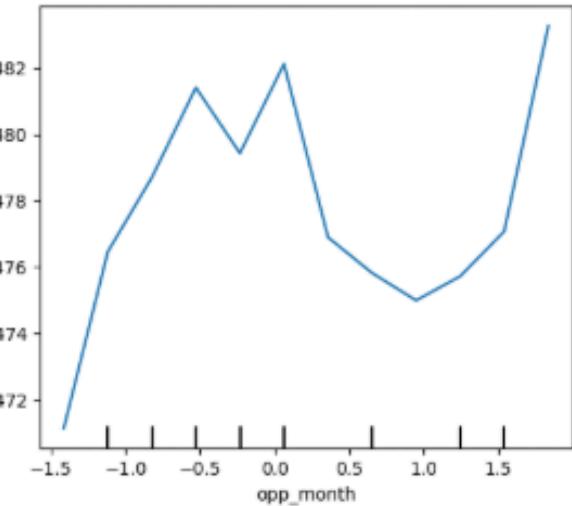
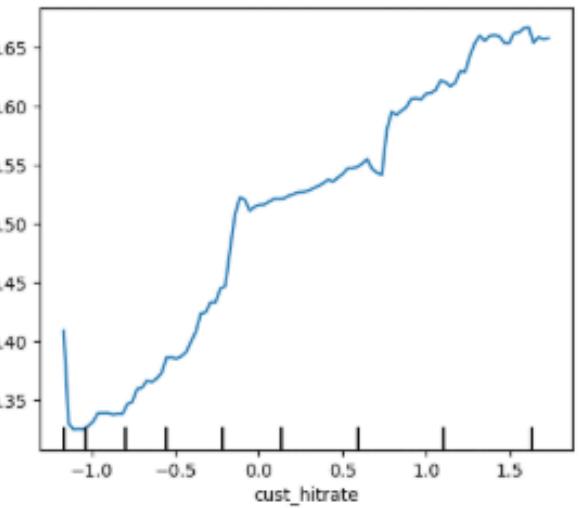
FLOP 3

Importancia de Características (KNN)

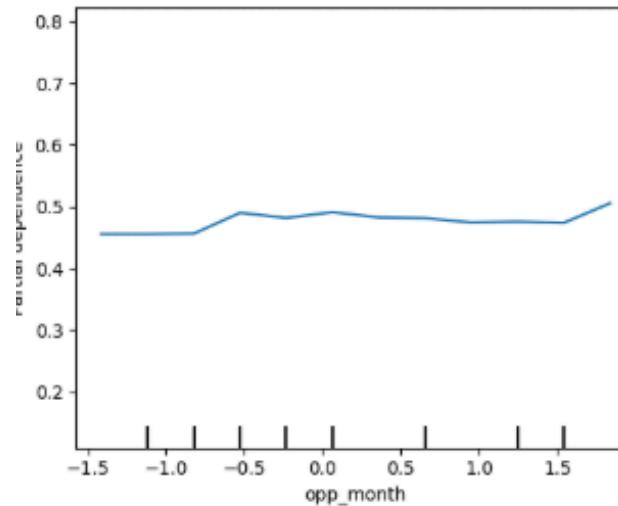
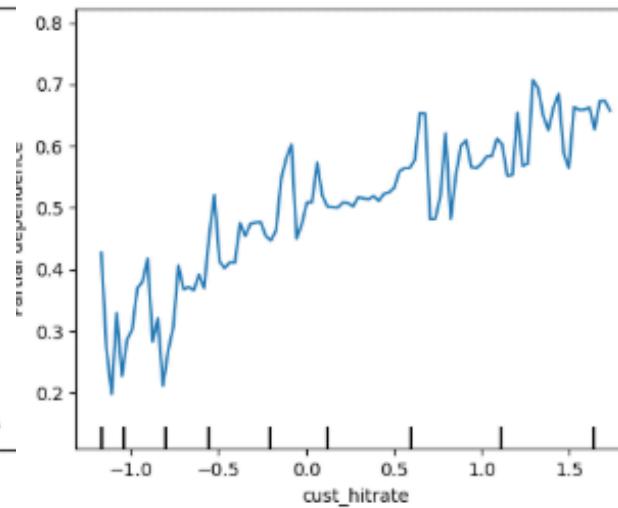


# Partial Dependence Plot (PDP)

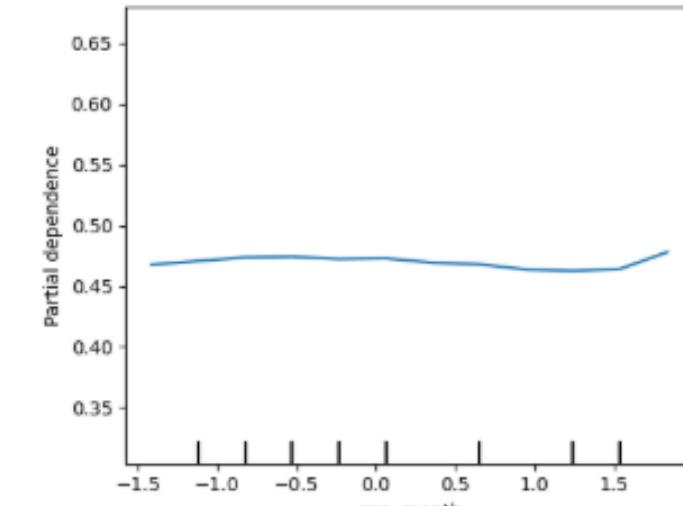
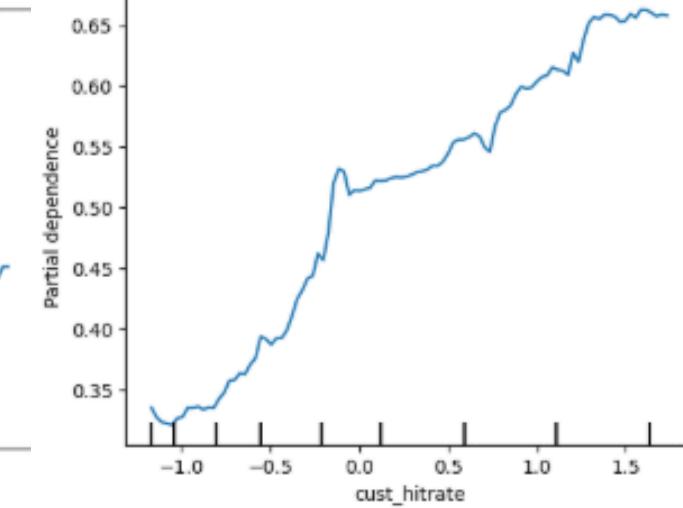
RANDOM FOREST



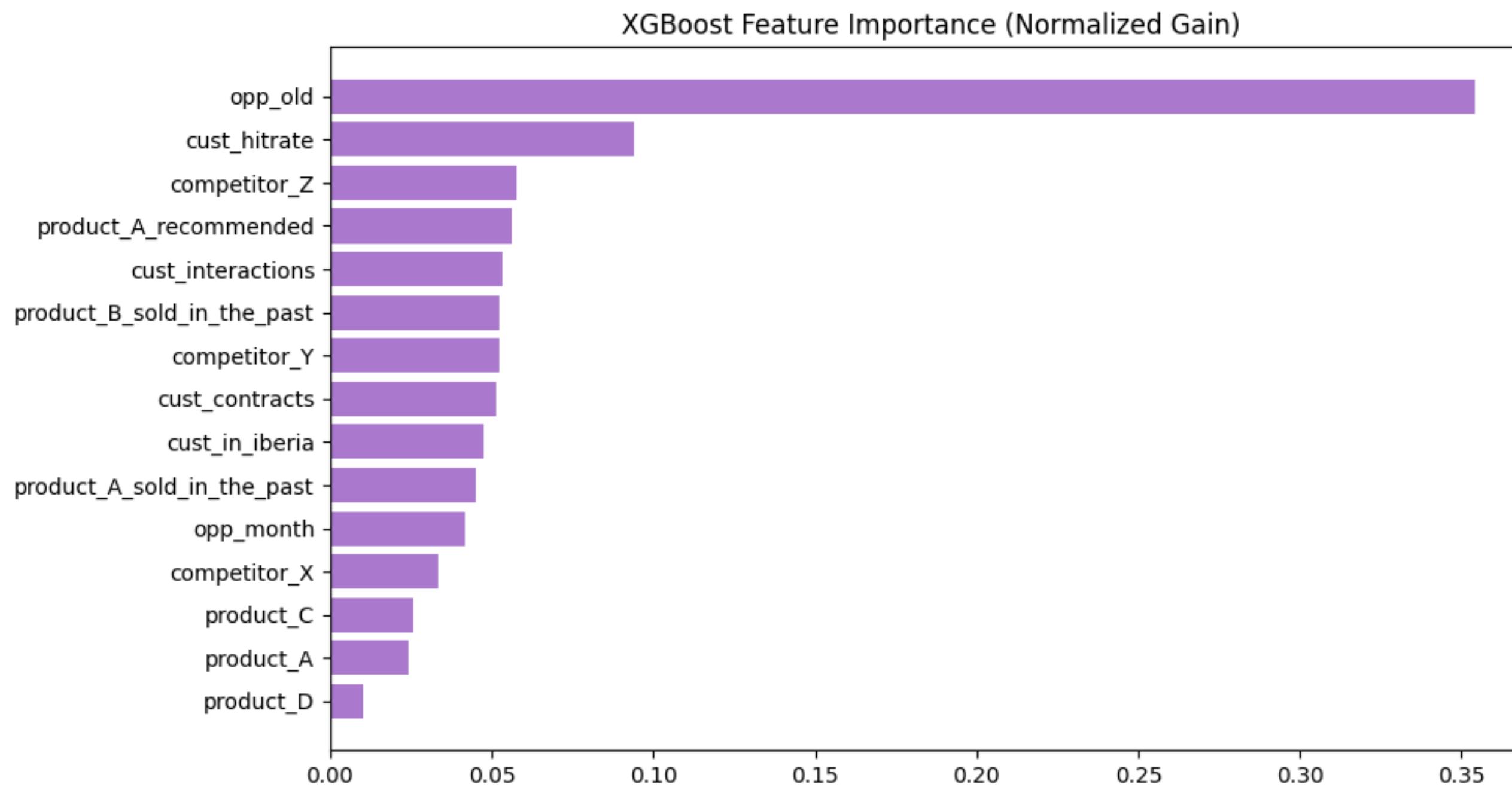
XGBOOST



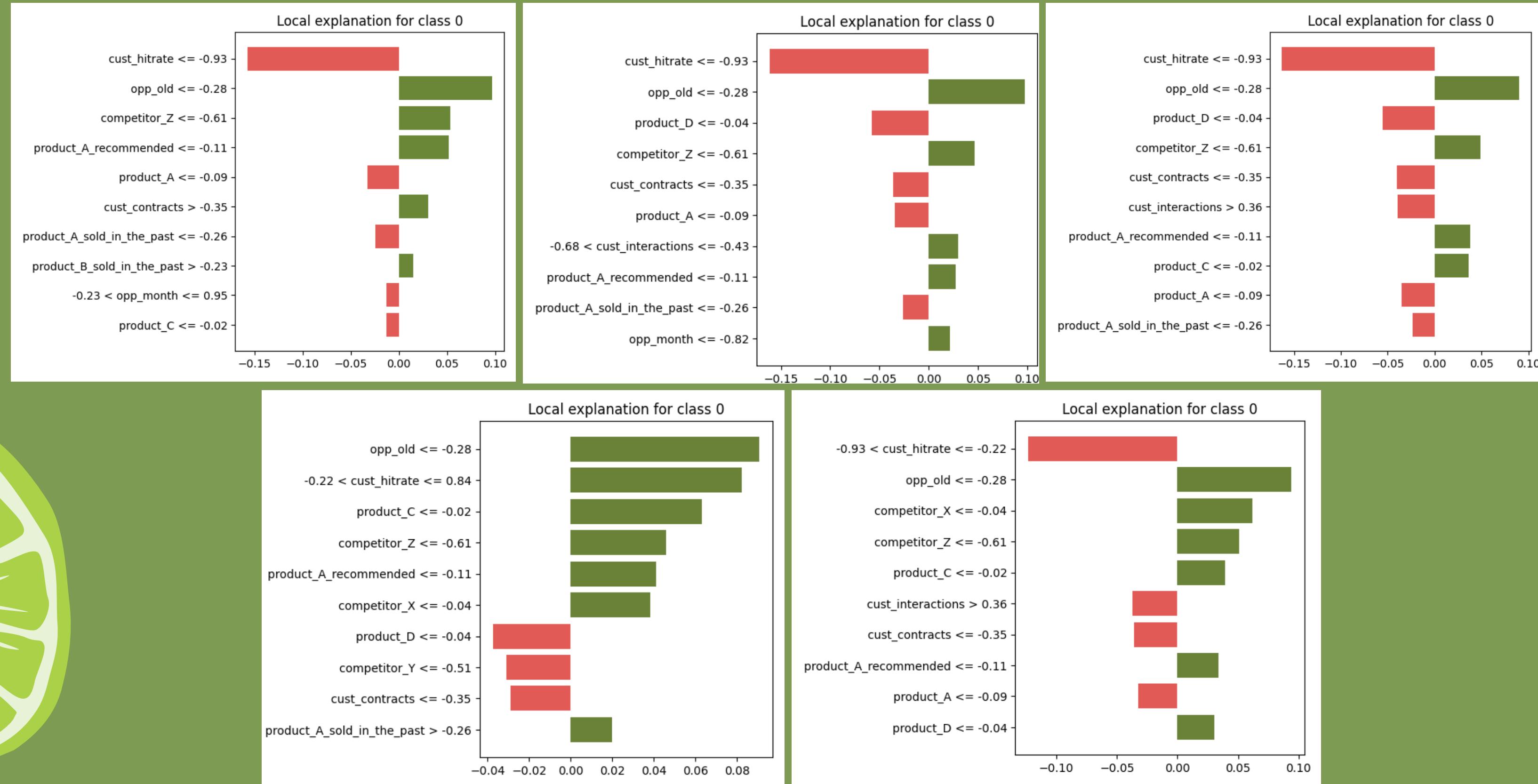
KNN



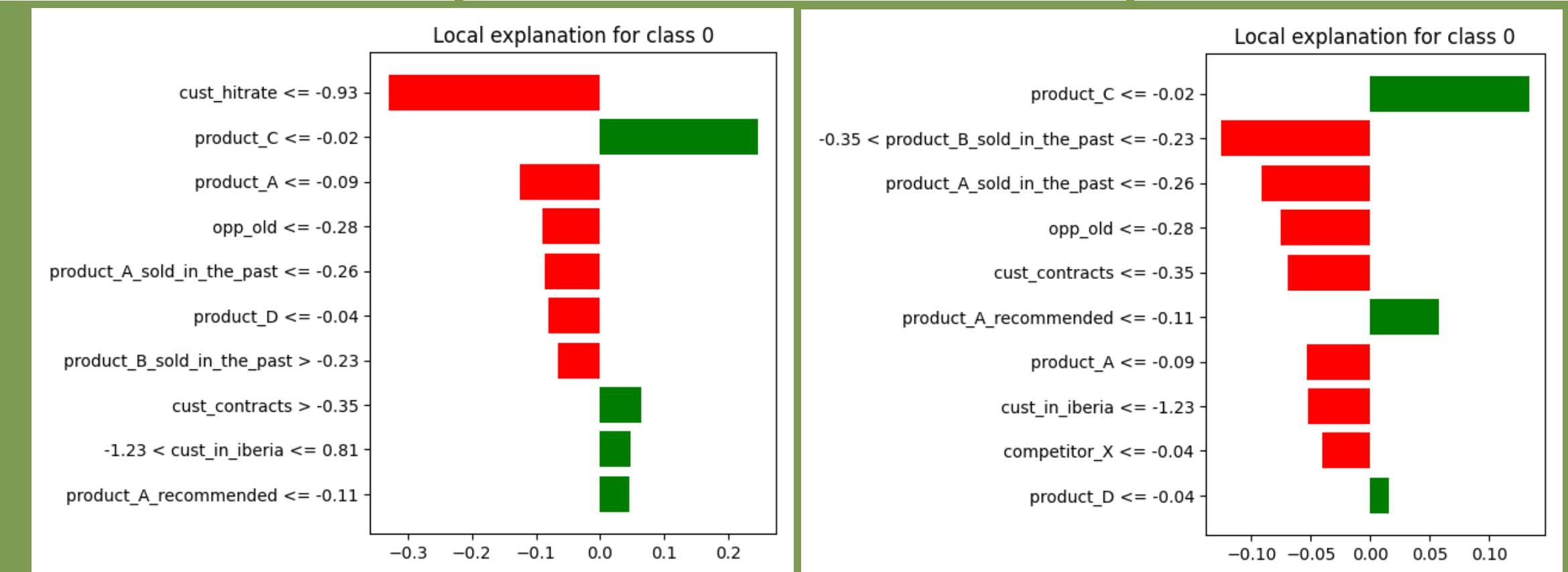
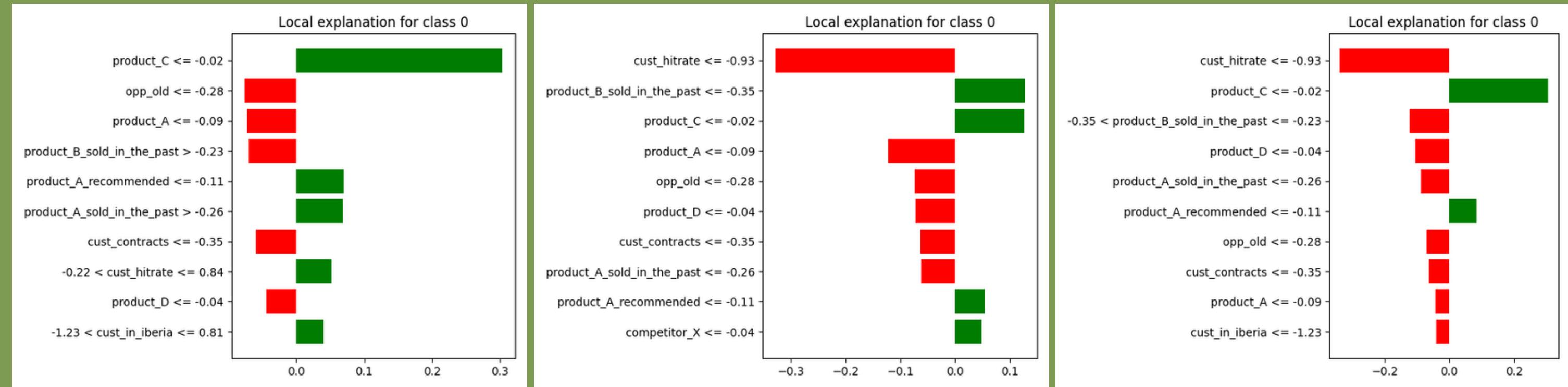
# XGBOOST GAINS



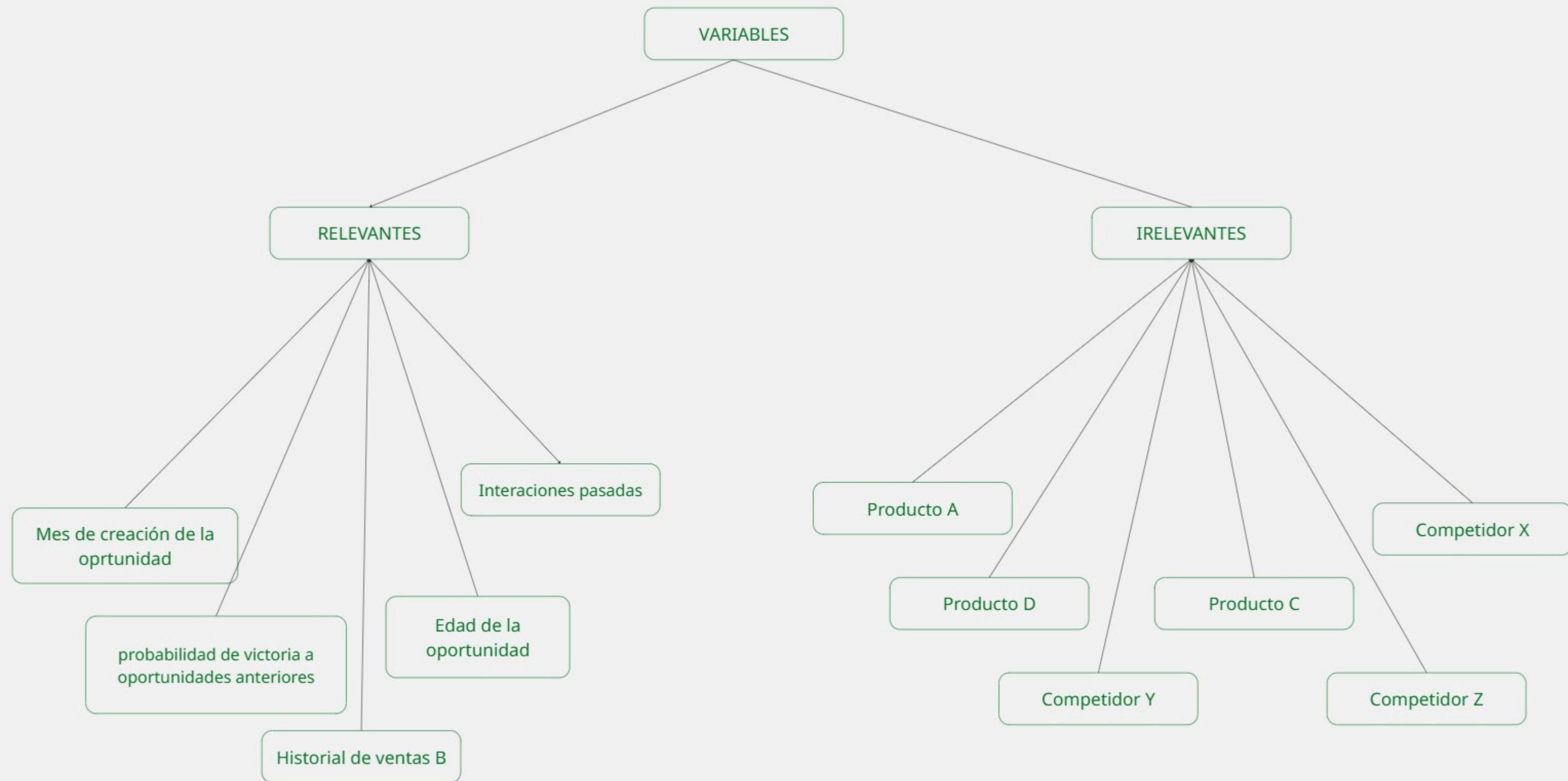
# LIME DE RANDOM FOREST



# LIME DE XGBOOST



# CONCLUSIONES



# ¿Qué haríamos si tuviéramos más tiempo?

- Combinar diferentes modelos de ML
- Aplicar más insights(global y local)
- Optimizar el código
- Buscar más métodos de análisis

# AGRADECIMIENTO

- Las identidades patrocinadoras y presentantes:  
**Schneider, Mango, Smadex, Qualcomm, Novaartis i Caixabank**
- El equipo de organizadores y mentores,y su inestimable labor
- La facultat de FME y ESTEIB por ofrecer sitio para el datathon

Mencions honorables:

- Nostre mentor de l'aula 103 , **Joel Márquez.**
- El autor del llibre adjunt a les instrucción del repte schneider **Christoph Molnar**
- Tots els **participants** que han gaudit del datathon

DANKE  
SCHÖN