

Análisis Predictivo 2025Q2

Examen 3

Gerónimo Fretes

1. Modelo base

1. Modelo base

- 396 Features
- Catboost
- Hiperparámetros predeterminados
- 1 fold sin validación
- Accuracy: 0.937

2. Selección de modelos

2. Selección de modelos

Modelo base modificado:

- 5 folds de 80% entrenamiento y 20% validación

En cada fold:

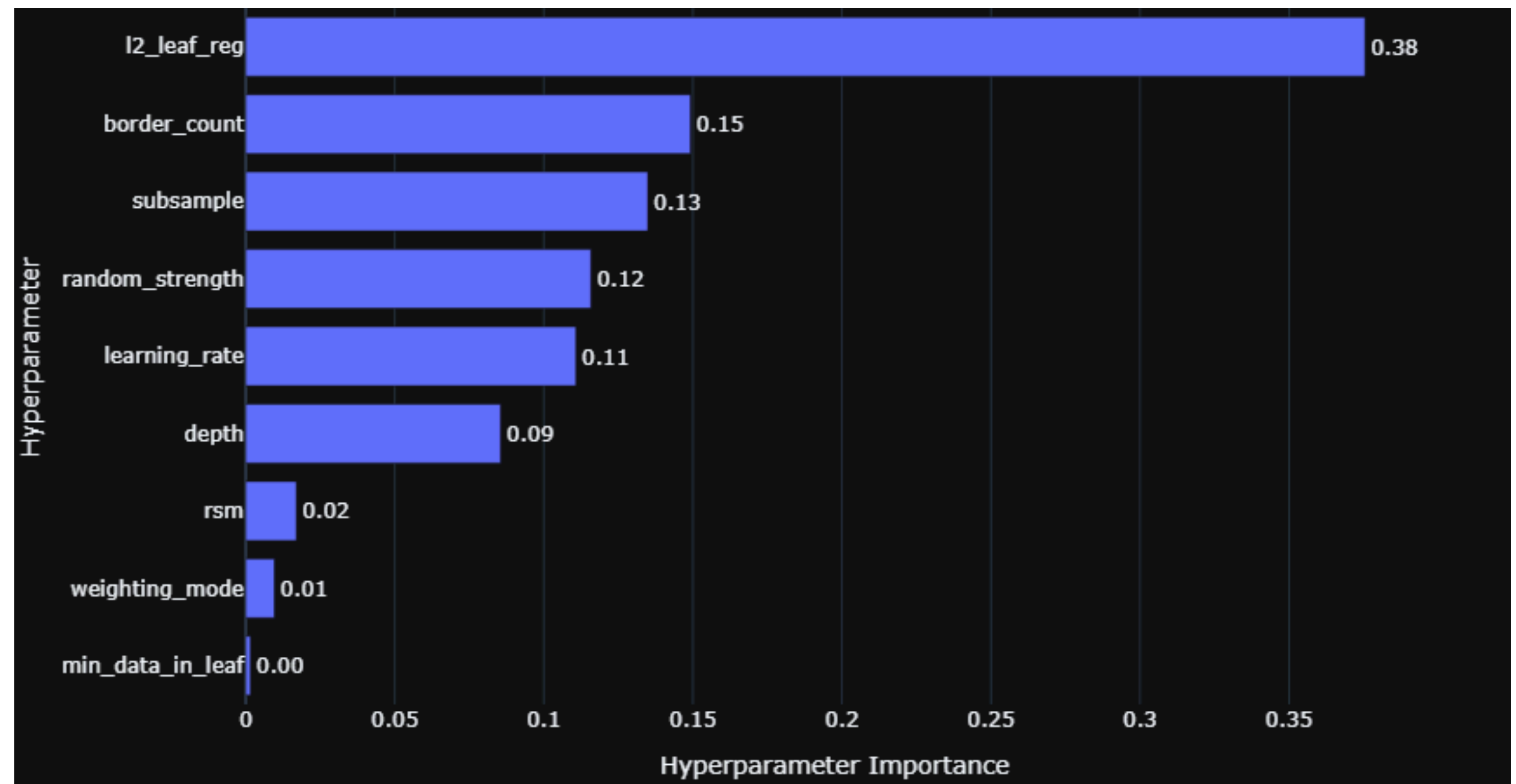
- Número alto de iteraciones máximas (5000)
- Detector de sobreajuste (deja de iterar luego de 200 iteraciones sin mejora en el ROC AUC)
- Selecciona la iteración que optimiza el accuracy en el split de validación
- Calcula el umbral de corte que optimiza el accuracy

2. Selección de modelos

Exploración de hiperparámetros con OPTUNA:

- Función a maximizar: accuracy promedio de los 5 folds

Hiperparámetros explorados:



3. Descripción del modelo final

3. Descripción del modelo final

Rendimiento:

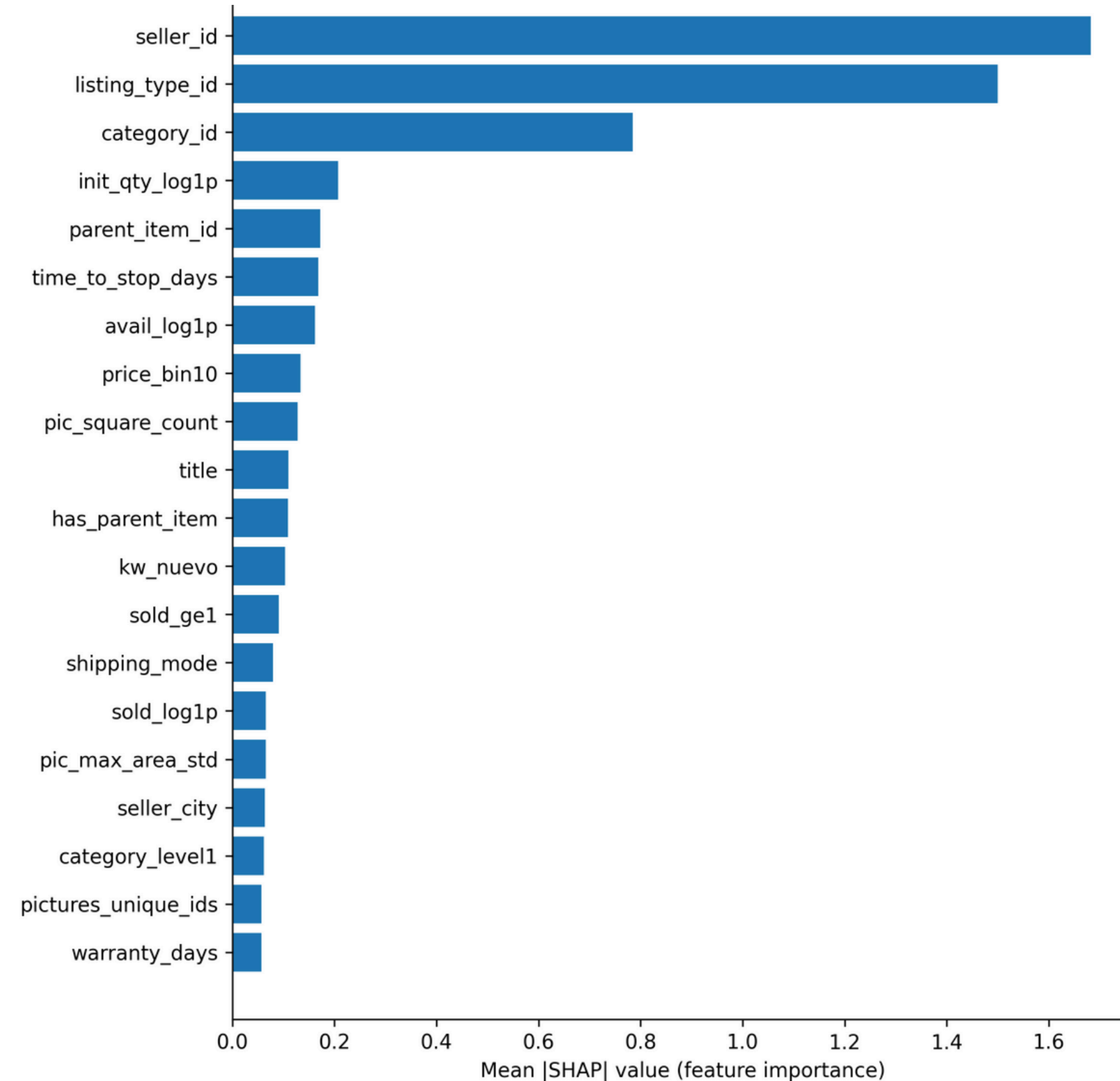
- Accuracy promedio en entrenamiento: 0.93704
- Accuracy en kaggle: 0.93744

Hiperparámetros seleccionados:

- depth: 7,
- l2_leaf_reg: 30.206492591943,
- learning_rate: 0.0226940853071826,
- random_strength: 5.70427449106795,
- min_data_in_leaf: 185,
- border_count: 121,
- subsample: 0.886336894885607,
- rsm: 0.922629694300833,
- weighting_mode: "auto_balanced"

3. Descripción del modelo final

SHAP value promedio en valor absoluto de los features más influyentes



4. Limitaciones y posibles mejoras

4. Limitaciones y posibles mejoras

- Aumentar la cantidad de intentos en la exploración de hiperparámetros
- Ampliar la búsqueda a distintos hiperparámetros
- Limpieza de features guiada por SHAP values

Respositorio