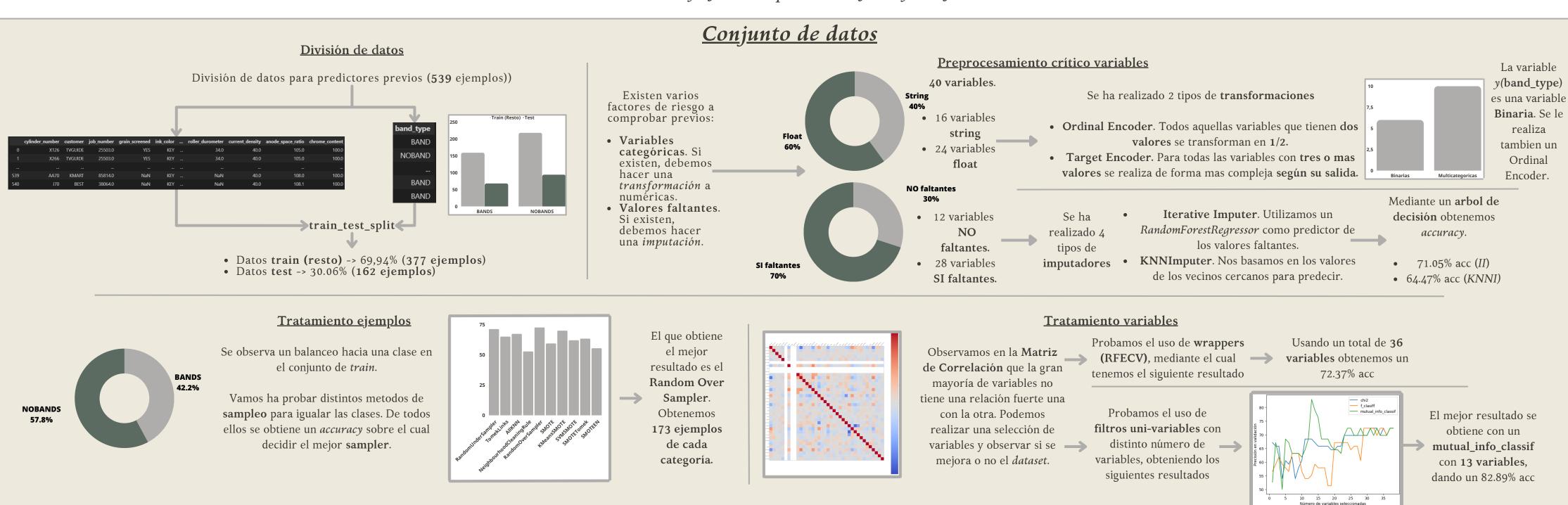
Julen Rodríguez Meneses Miguel Sagaseta de Ilúrdoz Sánchez

Predictor automatizado para la selección de bandas de cilindro

En la industria de la impresión, las bandas de cilindro son fundamentales para lograr una impresión de alta calidad y precisión. Seleccionarlas y verificarlas puede ser costoso y llevar tiempo, por lo que es importante desarrollar métodos automatizados y eficientes para identificar y verificar su calidad.

MOTIVACIÓN

Desarrollar un predictor eficiente para identificar bandas de cilindro en función de sus características físicas y de desempeño tiene el potencial de mejorar la calidad de la impresión y reducir los costos en la industria de la impresión.



Modelos evaluados

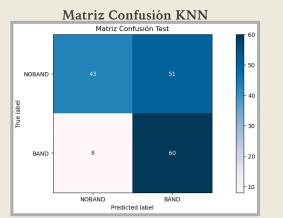
Decidimos utilizar los siguientes modelos para predecir resultados:

- **KNN**
- LogisticRegression

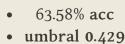
DecisionTreeClassifier

Por otro lado, se va ha utilizar como medida de elección de modelos el valor fscore.

A cada uno de ellos se le hará un ajuste de umbral de decisión para intentar mejorarlos.



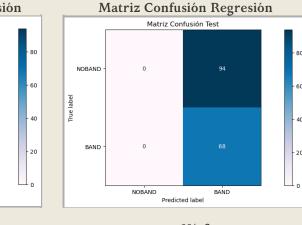
• 78.05% fscore





• 41.98% acc

• umbral 0.5

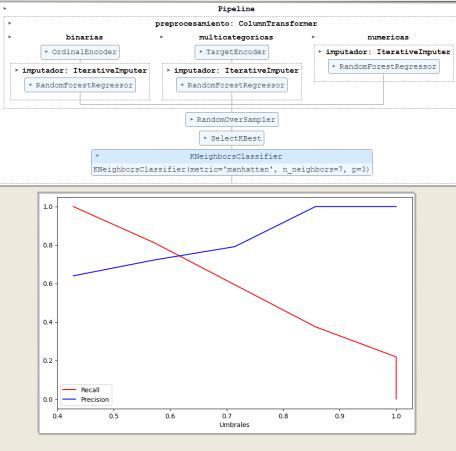


• 59.26% fscore • 41.98% acc

• umbral 0.5

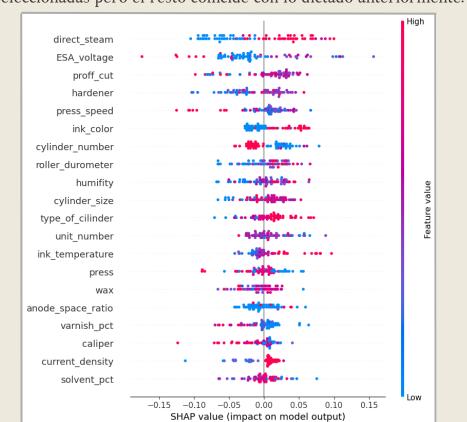
Estudio experimental

El mejor modelo sin duda el KNN. Si lo estudiamos tenemos los siguientes resultados.



<u>Interpretabilidad del modelo</u>

Si observamos la grafica basada en la **Permutation Importance**. Destacamos que la primera variable no fue una de las seleccionadas pero el resto coincide con lo dictado anteriormente.



Conclusiones y lineas futuras

A lo largo del proyecto nos hemos dejado varias ideas sobre el tintero. Además, muchos de los métodos que hemos aplicado no han proporcionado mejores resultados, con lo que no hemos podido estudiar su impacto real sobre nuestros datos en la predicción final.

En definitiva, como líneas futuras, nos gustaría mejorar nuestros métodos para probar más combinaciones de todos los

preprocesamientos que hemos probado. En concreto, nos gustaría estudiar el efecto de los outliers y del ruido sobre los datos, así como probar otras métricas de rendimiento.