

## Informe: Proyecto Final PRG.

Este proyecto tuvo como objetivo desarrollar un modelo de clasificación multiclase basado en características numéricas extraídas de un conjunto de datos astronómicos. Debido a que el archivo original no contenía una columna explícita de salida o “target”, se construyó una variable objetivo a partir de la columna `sy_num`, que representa el número de planetas en cada sistema estelar. Esta variable se utilizó para entrenar clasificadores capaces de predecir cuántos planetas conforman un sistema dado. La estrategia general se estructuró en varias etapas: primero se realizó una exploración de los datos numéricos, identificando posibles problemas de calidad y entendiendo la distribución de los atributos. Posteriormente, se aplicó un preprocesamiento que incluyó la imputación de valores faltantes con la media y la estandarización de todas las variables para asegurar una escala uniforme. A partir de allí, se entrenaron dos modelos: un `RandomForestClassifier` optimizado mediante `GridSearchCV`, y un `SVC` ajustado con `RandomizedSearchCV`. Ambos modelos se evaluaron usando métricas de clasificación y se compararon utilizando el coeficiente de Cohen’s kappa, que indicó una fuerte concordancia entre ellos. Los resultados mostraron que el modelo de Random Forest logró un mejor desempeño general, especialmente en clases menos frecuentes, siendo por tanto la opción recomendada en este caso.