Exoplanet Exploration

Summary

The objective of this project is to create machine learning models capable of classifying candidate exoplanets from the NASA Kepler space telescope dataset. 4 classification models were created: a logistic regression classifier, a k-nearest neighbors classifier, a random forest classifier, and a support vector classifier.

Of the 4 classifiers created, the random forest classifier performed the best with a f1-score of 0.89. The other classifiers performed similarly with a f1-score of around 0.80. The best performing random forest classifier may be good enough to predict new exoplanets since it had an average precision and recall of around 0.86. In the future, we could create a model that is better at predicting new exoplanets by using more data to train the model.

Classification Reports

Logistic Regression Classifier

	precision	recall	f1-score	support
CANDIDATE	0.66	0.51	0.58	422
CONFIRMED	0.62	0.72	0.67	450
FALSE POSITIVE	0.98	1.00	0.99	876
accuracy			0.81	1748
macro avg	0.75	0.74	0.74	1748
weighted avg	0.81	0.81	0.81	1748

Random Forest Classifier

	precision	recall	f1-score	support
CANDIDATE	0.82	0.74	0.78	422
CONFIRMED	0.78	0.82	0.80	450
FALSE POSITIVE	0.98	1.00	0.99	876
accuracy			0.89	1748
macro avg	0.86	0.85	0.85	1748
weighted avg	0.89	0.89	0.89	1748

K-Nearest Neighbors Classifier

	precision	recall	f1-score	support
CANDIDATE	0.63	0.50	0.56	422
CONFIRMED	0.60	0.70	0.65	450
FALSE POSITIVE	0.98	1.00	0.99	876
accuracy			0.80	1748
macro avg	0.74	0.73	0.73	1748
weighted avg	0.80	0.80	0.80	1748

Support Vector Classifier

	precision	recall	f1-score	support
CANDIDATE	0.69	0.47	0.56	422
CONFIRMED	0.61	0.78	0.68	450
FALSE POSITIVE	0.98	1.00	0.99	876
accuracy			0.81	1748
macro avg	0.76	0.75	0.74	1748
weighted avg	0.81	0.81	0.81	1748