The goal of this project was to take the NASA Kepler space telescope data and use it to process and predict scores.

The data file was preprocessed to remove unnecessary columns and null values. It was then scaled using the MinMaxScaler.

Once the data was scaled, it was split into training and testing groups.

Models:

SVC: Training Data Score: 0.8502592253735896

Testing Data Score: 0.838975297346752

Gridsearch:

"Use GridSearchCV to tune the C and gamma parameters"

grid.best_params_: {'C': 800, 'gamma': 1}

grid.best_score_: 0.8891430314120159

Changing the parameters slightly increased the score.

"Tune and compare at least two different classifiers."

Random Forest:

The first pass score was 89.3%. After researching the dataset, multiple columns were dropped due to lack of importance. Once this was done the model was re-run with a score of 89.57%.

KNN:

KNN proved to be worse than the original SVM model with a score of: k=5, Test Acc: 0.812

In conclusion, Random Forest is the best of the models used in this exercise with a 4.5% increase in score over SVC.