CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)

Krithi V. Koodli

Project Number

38376

Project Title

Distinguishing between Binary and Non-Binary Stars Using Machine Learning

Abstract

Objectives/Goals

My engineering goal was to classify stars as binary or non-binary stars using machine learning. When the classifier is given a feature, it should accurately label the star, and its predictions should have an accuracy of at least 85%. It should be able to predict the labels of stars which are yet unknown.

Methods/Materials

I found data sets from the Binary Star Database and Kepler Mission Stars to use for binary and non-binary stars, respectively. I chose the feature of orbital period, which is much smaller for binaries on a galactic scale, to train the model on because it was available or could be calculated for both data sets. Orbital period also distinguishes binaries from non-binary stars. Even though non-binary stars do not orbit and do not have an orbital period, using a hypothetical orbital period would allow the model to distinguish the two types of stars. I trained and tested both a KNeighbors Classifier and a Decision Tree Classifier and measured their cross-validation accuracies. After this, I tested it on only binary stars to measure its prediction accuracy. Subsequently, I tested it on stars whose labels were unknown. I used Python 2.7 with Sci-kit-learn to code the model.

Results

Using cross validation, I determined that the accuracy of the model was 97% and 91% for the KNeighbors and the Decision Tree, respectively. I then used the KNeighbors to predict on 4000 binary stars and determined its accuracy as 99%. The accuracy was sufficient, so I tested the model on data (using HYG database) where the stars' labels were unknown. Out of 200 tested, 51 were predicted to be binary stars. Through an independent verification, I have determined that at least two of the 51 stars are most likely binary.

Conclusions/Discussion

The model met the engineering goals and was able to accurately predict the label of stars. The model had an accuracy of 97% on both types of stars and an accuracy of 99% on only binary stars. The model was able to predict the label of unclassified stars, and two of the 51 predicted stars are likely binary stars. The final algorithm had a sufficient accuracy to be implemented in databases to scan for binary systems. This expands our knowledge of physics and astronomy by providing a novel method to distinguish binary and non-binary stars using the orbital period, and to find overlooked binary stars. I will continue to verify the binary stars predicted by the model.

Summary Statement

My project created a model that classified stars as binary stars or non-binary stars by training and testing on the orbital period feature, and it predicted the label of unclassified stars in stellar databases by using machine learning.

Help Received

I received help from my brother, who helped me troubleshoot some of the errors I received in the process of creating the model