Machine Learning Project Abstract

Jacob Matteo and Nico Veneziano

https://www.kaggle.com/keplersmachines/kepler-labelled-time-series-data

1. **What is the problem?**

   a. The Kepler space telescope was designed to be very sensitive to light in order to possibly detect planets using the transient method. This method of finding planets works by watching stars to see if they dim and brighten at regular intervals. If a regular dimming happens, that means a planet passed infront of the star. During the lifetime of the telescope, Kepler took data from many stars yet to be sorted through. One of the campaigns, specifically Campaign 3, was shown to have no planets, but given the amount of stars, the likelihood of this is slim so many are skeptical of this result. The goal is to confirm the number of stars looked at during this campaign that have planets orbiting around them.

2. **How are people currently solving the problem? What is currently state of the art?**

   a. Currently, the data gets automatically looked over by a basic program which detects light pulses and then positives from the simple algorithm are further refined by people.

3. **Why do current solutions not work?**

   a. While this approach may get a lot of results, ai tends to be able to find patterns in data that are not as easily spotted by more simplistic algorithms and even people. There may be many stars with planets missed in the data set. A small number of ai's have been used to double-check this data but additional results are few and far between.

4. **What and how do you propose to solve the problem?**

   a. By creating a deep learning model to perform the exoplanet hunting, we will be able to train the model using the training data in order to use it on the testing data.

The goal will be to use machine learning techniques to find the relationship between the change in flux and whether the star does have an exoplanet paired with it or not. However, the imbalance of data (in terms of stars which have exoplanets present and stars which do not) will have to be addressed first.

5. **What is your plan of action?**

   a. Our plan of action will be to analyze the dataset in order to find the best deep learning model to fit our data to. Data preprocessing will have to occur first to ensure that the data will be optimized for the model itself. This model will be created and trained on the data then tested and scored through loss and accuracy metrics. The model will be optimized and adjusted in order to find the most optimized architecture to achieve the highest accuracy while avoiding overfitting.