Machine Learning – Kepler exoplanet project

Data: Downloaded from <https://www.kaggle.com/nasa/kepler-exoplanet-search-results>.

Data Cleaning: Data cleaning done in several steps. The columns that contained names were dropped ('rowid','kepid','kepoi\_name','kepler\_name','koi\_pdisposition', 'koi\_score' ,'koi\_tce\_delivname'). 'koi\_pdisposition’ was also dropped dropped because this result is what we plan to predict using our model. Also the two columns 'koi\_teq\_err1','koi\_teq\_err2' were dropped because they were all null values. In addition to that rows with null vales were dropped. Also rows with 'koi\_pdisposition’ = ‘CANDIDATE’ were dropped because they were not useful in training the model. The original data set had9564 rows and the cleaned data set contains 6031 rows. The cleaned data set was saved in ‘exoplanet\_cleaned.csv’.

Preprocessing Data: Data was first separated to X and y values. 'koi\_pdisposition’ is used as the y values. X has 40 input variables. X and y were reshaped and converted to Numpy arrays. The data was then separated to training(75%) and testing(25%) samples. The label vales ‘CONFIRMED” and ‘FALSE POSITIVE’ in y were encoded (label and one hot encoding) for the Neural network.

Comparison of Models:

Initially tried the Models logistics and SVM.

The results are as follows:

Logistics:

A screenshot of a cell phone

Description automatically generated

SVM:

A screenshot of a cell phone

Description automatically generated

Neural Network: Loss: 0.4747426787253672, Accuracy: 0.9909529685974121

Both models logistics and SVM performed similarly. I have chosen the SVM model as my best model. The simple neural network I created has high accuracy (99.1%) but the loss is 47.5%. It is much slower compared to other two models.

File Names:

cumulative.csv – original data file

cumulative\_cleaned.csv – cleaned data file

exoplanet.ipynb – notebook file to clean data

logistic.ipynb – notebook file for logistic model

Machine Learning.docx – this document

neuralnet.ipynb - notebook file for neural network file

SVM.ipynb - notebook file for SVM model

SVM\_model.sav – saved model