RESEARCH INTERNSHIP
PROGRESS REPORT

Project Title: Classification for Potentially Hazardous Asteroids
Team Member's Name: Jishnu N
Guide: Dr. C. Kavitha
Internship Duration: 14 November, 2022 – 11 January, 2023
Summary: The main focus of this project was to use multiple supervised learning classification algorithms for the identification of Potentially Hazardous Asteroids (PHAs) through asteroid characteristics. Logistic Regression, K- Nearest Neighbours, Support Vector Machine, Random Forest, XGBoost, Balanced Bagging were used in this study.

Week	Start Date	End Date	Objectives Accomplished
Week 1	14/11/2022	21/11/2022	Investigated potential datasets for the project. Dataset from Nasa API was considered.
Week 2	21/11/2022	28/11/2022	Find out the relevant attributes needed for the problem statement in the project and understand what each feature in the data set signify.
Week 3	28/11/2022	05/12/2022	Examined and learned about various pre-processing techniques. Employed feature reduction by determining their inter-correlation.
Week 4	05/12/2022	12/12/2022	Researched about outliers and their influence in the data set. Performed scaling of outliers, feature encoding and feature scaling.
Week 5	12/12/2022	19/12/2022	Difference between linear vs non-linear algorithms were evaluated and understood which algorithms were best suited for the project. Decided which algorithms each team member would work on separately. I had selected Random forest and XGBoost classifier algorithms to work on. Distribution plots and Gini index was calculated.
Week 6	19/12/2022	26/12/2022	Extensive study on random forest algorithm was done. Model creation, model training, model evaluation and hyperparameter tuning was done to obtain predictions. Model's performance was evaluated using confusion matrix, accuracy score metric, precision score metric, recall score metric, and the F1-score metric. Hyperparameter dependence were studied using plots to perform optimum tuning and to check for underfitting and overfitting problems.
Week 7	26/12/2022	02/01/2023	Extensive study on XGBoost algorithm was done. Model creation, model training and model evaluation was done to obtain predictions. Model's performance was evaluated using confusion matrix, accuracy score metric, precision score metric, recall score metric, and the F1-score metric. Feature importance was plotted.
Week 8	02/01/2023	11/01/2023	Collectively prepared the internship report with the teammates.