Project Summary

We developed multiple machine learning models trained on NASA's open exoplanet datasets to automatically identify potential exoplanets from large light curve datasets. We'd also developed our own data cleaning software to merge the Tess and Kepler datasets for a model used on that. Given more time, it would be possible to automate this data cleaning program so that it would be compatible with almost all datasets within Nasa's database. Of the three different models, we actually had one model reach 92% accuracy after tuning. This project directly addresses the challenge of manual exoplanet identification by automating and enhancing accuracy in data analysis. It's important because it accelerates discovery, supports NASA's mission of exploring habitable worlds, and showcases how AI can advance astrophysical research globally. Please do bear in mind however, that although we developed three models, only one has been fully tuned due to time constraints. We were also unsuccessful in completing the front-end side of development because of time constraints. The back-end is posted and ready to be looked at though.

 What it does:
Automatically detects potential exoplanets from NASA's light curve data using multiple machine learning models.

 How it works:
We have created 3 models, each with their own hyperparameters which were determined using extensive validation testing. We used random forest for all the models. The hyperparameters were then used to train the most accurate model on the training dataset, which was tested against the test set to determine the accuracies. Interestingly, the models differ only in hyperparameters and datasets. These datasets included the Kepler dataset, the Tess dataset, and then finally a model for the combined dataset created by the cleaner.

Benefits:

- Speeds up exoplanet discovery

- Reduces human error and manual analysis

- Provides high detection accuracy (92%)

- Scale for future NASA missions

Goals:
Enhance automated detection accuracy, expand on the data cleaning program, and make the tool open source for global researchers.

Tools & Technologies:
Python, ChatGPT, Gemini, Nasa Open Databases.
 Use of Artificial Intelligence

We used Gemini, and ChatGPT to understand where we did wrong with the coding and took help from it to understand and fix the errors and also we used it as a learning tool.