PROJECT PROPOSAL

Enhancing Exoplanet Detection Accuracy through Machine Learning Models.

Introduction:

Space is an area that fascinates me the most. Exoplanets are the planets that orbits a star outside our solar system. We can use space telescopes like Kepler for the detection of exoplanets. It detects the exoplanets analyzing the light curves to identify potential transits of planets across their host stars. While this method is practically used and is effective, it does generate a number of False Positives from other astrophysical phenomena. The goal of this project is to apply Advanced Machine Learning techniques to data collected from Kepler telescope to improve the accuracy and reliability of the exoplanet detection.

Objective and Dataset:

The main objective is to build a machine learning model and tune it so that it can distinguish between true exoplanets and false positives more effectively. The dataset that I used is "Exoplanet Detection on Kepler Data" from Kaggle, which contains labeled instances of confirmed exoplanets and false positives, with a variety of features derived from the light curves.

Dataset: Exoplanet Detection on Kepler Data

Methodology:

1. Data Preprocessing: Standardizing the dataset, managing missing values, and exploring data transformations to enhance model performance.
2. Feature Engineering: Creating new features that capture additional information from the light curves, such as statistical descriptors or transformations that may highlight aspects indicative of exoplanetary transits. [3]
3. Model Development: Build a baseline with logistic regression to understand the basic performance and set a benchmark for improvement. Implementing ensemble methods such as Random Forests and Gradient Boosting Machines (GBM) to leverage their robustness against overfitting and ability to handle imbalanced data.[2] Exploring deep learning models (if time permits), particularly convolutional neural networks (CNNs), which can directly process light curves and might capture temporal patterns indicative of exoplanetary transits.[1]
4. Model Evaluation: Utilizing k-fold cross-validation to assess models and applying techniques such as grid search for hyperparameter tuning. Performance metrics will include accuracy, precision, recall, F1-score, and ROC-AUC.

References:

1. Shallue, C. J., & Vanderburg, A. (2018). Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *The Astronomical Journal*, 155(2), 94. [Deep learning approaches and their effectiveness in exoplanet detection].

2. Armstrong, D. J., et al. (2017). Automatic vetting of planet candidates from ground-based surveys: Machine learning with NGTS. *Monthly Notices of the Royal Astronomical Society*, 471(3), 2801-2812. [Use of ensemble methods in exoplanet detection].

3. Ansdell, M., et al. (2018). Scientific Domain Knowledge Improves Exoplanet Transit Classification with Deep Learning. *The Astrophysical Journal Letters*, 869(1), L7. [Importance of feature engineering in improving model performance].