Peter Santana
Final Report Proposal

AI395T
University of Texas at Austin

Predictive Assessment of Near-Earth Comets' Hazard Potential Using Machine Learning

Introduction:
Near-Earth comets are celestial bodies with orbits that are close to Earth's orbit. Understanding their trajectories and predicting potential hazards are critical for planetary defense. Traditional methods of assessing collision risks involve extensive calculations and simulations, which can be enhanced by applying machine learning techniques to automate and improve the accuracy of predictions. This project aims to develop a machine learning model that predicts the hazard potential of near-Earth comets based on their orbital elements, such as eccentricity, inclination, and Minimum Orbit Intersection Distance (MOID). Methods similar to the one that wants to be accomplished by this project has been used and proved with near-earth asteroids as well.

Objectives:

1. To explore and analyze the orbital parameters of near-Earth comets and understand their significance in determining potential hazards.
2. To develop a robust machine learning model that can predict the hazard potential of comets based on their orbital elements.
3. To evaluate the performance of the model using appropriate metrics and refine it to improve prediction accuracy.
4. To interpret the model's results and identify the key features contributing to the hazard assessment, aiding in future astronomical studies.

Methodology:

The dataset is from NASA's Open Data Portal, which contains over 170 NECs (near earth comets) sourced by NASA's Jet Propulsion Laboratory's Near Earth Object Program.

Model Development:
- Machine learning models, including Random Forest, Support Vector Machines, and Logistic Regression, will be trained to classify comets as "hazardous" or "non-hazardous."
- Hyperparameter tuning will be conducted to optimize model performance.

Model Evaluation:
- Evaluation metrics such as accuracy, precision, recall, and F1-score will be used.
- Cross-validation will ensure the model generalizes well to new data.

Interpretation and Visualization:
- Feature importance analysis and visualization techniques (e.g., SHAP plots) will be used to interpret which orbital parameters are most indicative of hazard potential.

Expected Outcomes:
A machine learning model that accurately classifies near-Earth comets based on their orbital elements, providing reliable predictions of hazard potential.