A Study on Asteroid Characteristics and Impact Assessment Using Machine Learning

Sana S. Valappil

Abstract

Understanding the characteristics of asteroids and their potential impact on Earth is essential for planetary defense and space exploration. This article explores the use of machine learning techniques to evaluate asteroid data and determine the relationship between an asteroid's properties and its hazard level. The random forest model is used to identify the most important features for predicting the hazardous nature of an asteroid.

Introduction

Asteroids are little, rocky objects that orbit the Sun in the asteroid belt. These celestial bodies are considered to be remnants of the early solar system, and they are both scientifically interesting and potentially dangerous to Earth. Although most asteroids maintain stable orbits, others are pushed along trajectories that bring them dangerously close to Earth. These are called Near-Earth Objects (NEOs), and they are of particular concern since they have the potential to collide, causing serious damage (Perna et al., 2013). Historical evidence, such as the impact that caused the extinction of dinosaurs 66 million years ago, emphasizes the significance of monitoring and assessing these hazards. This article examines asteroid data, which includes the magnitude, estimated diameter, close approach date, relative velocity, and various orbital characteristics, using the random forest model.

Methodology

The dataset provided by NASA's Center for Near Earth Object Studies was used to conduct this study. The asteroid dataset consisted of 40 features that included the asteroids' Neo Reference IDs, their names, and several physical and orbital features describing them. For the initial data cleaning process, the following steps were taken. First, features such as the Neo Reference ID, the name, and the orbiting body were removed as it does not affect the nature of the asteroid. Additionally, there were redundant features with different units of measurement, such as estimated diameter in kilometers, meters, miles, and feet, as well as relative velocity in kilometers per hour and miles per hour. To ensure consistency, all the measurements with kilometers was chosen. Finally, the date features, close approach date and orbit determination date, were distributed into its year, month, and day components to enhance usability. After the data cleaning process, we reduced the dataset into 29 features. The dataset had no missing values or duplications. It is important to note that the dataset is unbalanced, so stratification was used while splitting the data into training and testing.

For feature selection, the features with near zero variance were found. The minimum orbit intersection was found to have very little variance, so it was removed. Additionally, the features were normalized and the correlation between the features was computed. According to the correlation heatmap in Figure1, the year of the close approach, the Jupiter Tisserand Invariant, the perihelion time, the semi major axis, the orbital period, and the mean motion were the features that were removed.

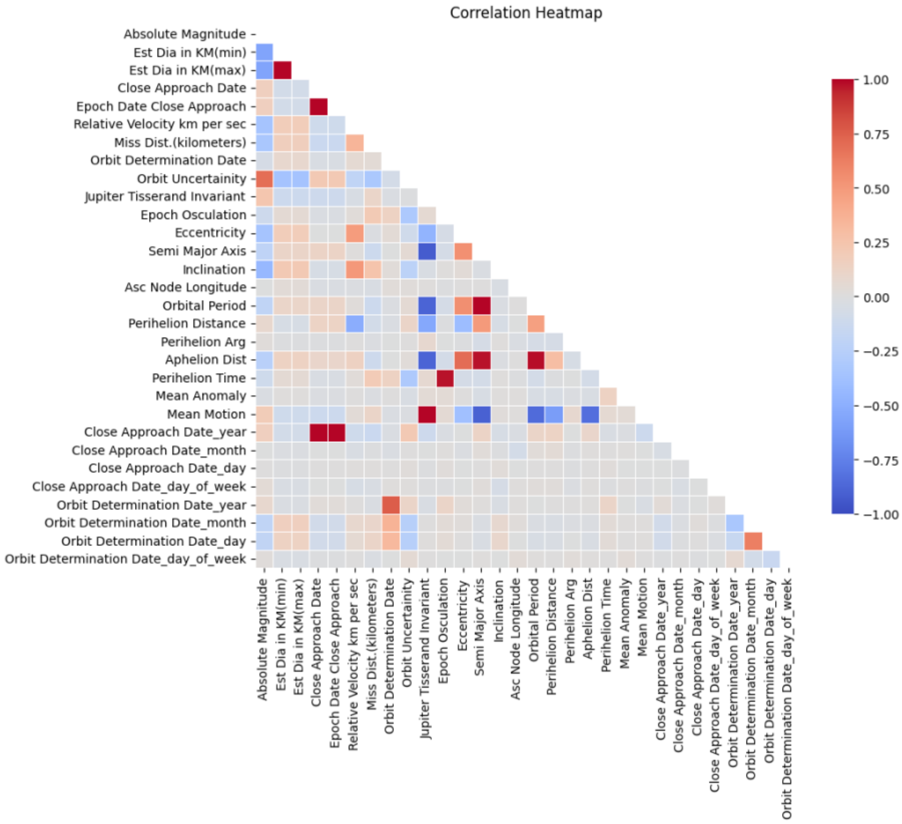

Figure 1: Correlation heatmap of the features of the asteroid dataset

Results

The random forest model was applied on the final 22 features that were chosen. The model was found to have an accuracy of 85%. Based on the random forest model, the order of the significant features was found. As seen in Figure2, the inclination of the asteroid is the most significant feature, suggesting that the tilt of an asteroid's path relative to the Earth's orbital plane has a strong correlation with its potential to be hazardous. This may be due to higher inclinations increasing the chance of intersecting Earth's orbit under certain conditions. Followed by the perihelion distance, absolute magnitude, indicating that both proximity to Earth's orbit and the size/brightness of the object are key indicators of potential threat. The minimum and maximum estimated diameter of the asteroid also ranked highly. Other orbital parameters like aphelion distance (farthest point from the Sun) and the argument of perihelion (which affects how close the asteroid gets to Earth when crossing the orbit) also showed strong influence, reinforcing the importance of orbital dynamics in threat assessment.

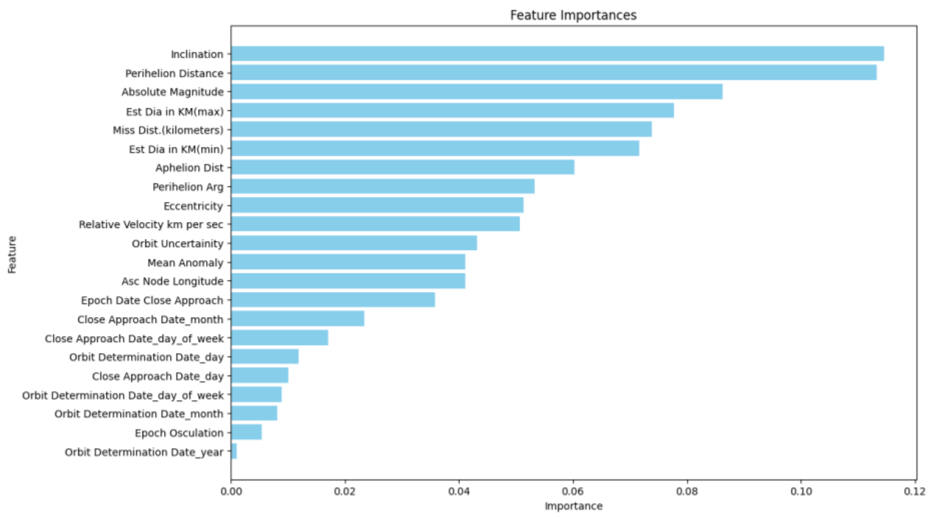

Figure 2: Feature importance based on the random forest model

Conclusion

Machine learning-based research on asteroids provides potential insights into their features and impact risks. By using new analytical methods, we can better detect and minimize possible asteroids risks while also learning more about these ancient remnants of the solar system. Future studies can employ different machine learning techniques to include whether the significant features are positively or negatively correlated to the hazardous nature of an asteroid.

Acknowledgements

This study was conducted with the NASA dataset that was uploaded on Kaggle by Lovish Bansal.

References

Bansal, L. (2024, March 27). *NASA: Asteroids classification*. Kaggle. https://www.kaggle.com/datasets/lovishbansal123/nasa-asteroids-classification

Perna, D., Barucci, M.A. & Fulchignoni, M. The near-Earth objects and their potential threat to our planet. *Astron Astrophys Rev* **21**, 65 (2013). https://doi.org/10.1007/s00159-013-0065-4