Team 24: Identifying Hazardous Near-Earth Asteroids Using Machine Learning

1 Introduction

The `24.csv` **Near-Earth Object Web Service (NeoWs)** dataset provides information about asteroids that closely approach Earth. This dataset is derived from NASA's **Near Earth Object (NEO)** database and includes key attributes related to an asteroid's physical properties, orbital parameters, and potential threat level.

The primary objectives of this dataset are to:

- **Classify asteroids** as hazardous or non-hazardous.
- **Identify key features** that contribute to an asteroid being classified as potentially hazardous.

This dataset enables the development of machine learning models to enhance asteroid threat detection, supporting astronomical research and planetary defense strategies.

2 Dataset Description

The dataset consists of various attributes related to near-Earth asteroids, including:

2.1 Orbital and Physical Attributes

- **Asteroid ID** – Unique identifier for the asteroid.
- **Close Approach Date** – The date when the asteroid is closest to Earth.
- **Absolute Magnitude** – Measure of the asteroid's brightness.
- **Estimated Diameter** – Minimum and maximum estimated diameter of the asteroid.
- **Relative Velocity** – Speed of the asteroid relative to Earth (km/s).
- **Miss Distance** – Distance between the asteroid and Earth at closest approach (km, lunar distances, astronomical units).
- **Orbiting Body** – Celestial body the asteroid orbits (e.g., Earth, Sun).
- **Orbital Eccentricity** – Measure of how elongated the asteroid's orbit is.
- **Semi-Major Axis** – Longest radius of the asteroid's elliptical orbit.
- **Inclination** – Angle between the asteroid's orbital plane and Earth's orbital plane.

2.2 Target Variable (Classification Task)

- **Potentially Hazardous Asteroid (PHA)** – Binary classification:
  - **Hazardous** – The asteroid poses a potential threat to Earth.
  - **Non-Hazardous** – The asteroid is not considered a threat.

3 Tasks and Requirements

To develop a machine learning model for asteroid classification, the following tasks need to be performed:

3.1 Data Preprocessing and Feature Engineering

- Handle missing values and clean inconsistent records.
- Normalize numerical attributes such as velocity, miss distance, and orbital parameters.
- Encode categorical variables (e.g., orbiting body) for model compatibility.
- Perform feature selection to identify the most relevant asteroid characteristics.

3.2 Hazardous Asteroid Classification (Supervised Learning)

- Train classification models.
- Evaluate model performance using **accuracy, precision, recall, F1-score, and ROC-AUC**.
- Compare multiple models to determine the most effective approach.
- Identify the most influential features that contribute to an asteroid being classified as hazardous.

3.3 Visualization and Reporting

- Generate bar charts and scatter plots to analyze asteroid distributions.
- Create correlation heatmaps to explore relationships between asteroid features and hazard classification.
- Use confusion matrices to assess classification performance.

4 Submission Requirements

- A well-structured report detailing the methodology, results, and analysis in a given report format.
- Python code is used for implementation.
- A presentation summarizing key findings and recommendations in a given presentation format.