**Project Report: Asteroid Hazard Prediction -Made by Asteroid Analyzers**

**1. Introduction**

The objective of this project was to develop a machine learning model to predict whether an asteroid is hazardous based on various features. The model utilizes data from asteroids, including characteristics such as orbital parameters, size, and reflectivity.

**2. Data Overview**

**2.1 Training Data**

* **File:** train.csv
* **Columns:**
  + name: Identifier for the asteroid (not used for prediction)
  + a, e, i, om, w, q, ad, per\_y, data\_arc: Orbital and physical parameters
  + condition\_code, H, neo, pha: Additional attributes, including whether the asteroid is classified as a Near-Earth Object (NEO) and Potentially Hazardous Asteroid (PHA)
  + diameter, albedo, rot\_per, moid: Additional features related to the asteroid's size and rotation

**2.2 Test Data**

* **File:** test.csv
* **Columns:** Same as the training data but without the target variable neo.

**3. Preprocessing**

**3.1 Handling Missing Values**

* **Numerical Columns:** Filled missing values with the mean of the respective column.
* **Categorical Columns:** Filled missing values with the mode (most frequent value) of the respective column.
* **Categorical Encoding:** Converted neo and pha columns to binary values (Y=1, N=0).

**3.2 Feature Selection**

* Dropped columns that were not relevant for prediction (name, condition\_code, pha).
* Removed highly correlated features based on a correlation threshold of 0.9 to avoid multicollinearity.

**3.3 Data Splitting**

* **Training Set:** Used to train the model.
* **Validation Set:** Used to evaluate the model during training.

**4. Model Development**

**4.1 Model Architecture**

* **Type:** Neural Network
* **Layers:**
  + Input layer with 64 neurons and ReLU activation
  + Dropout layer (0.5 dropout rate)
  + Dense layer with 32 neurons and ReLU activation
  + Dropout layer (0.5 dropout rate)
  + Output layer with 1 neuron and sigmoid activation

**4.2 Compilation and Training**

* **Optimizer:** Adam
* **Loss Function:** Binary Crossentropy
* **Metrics:** Accuracy
* **Training Parameters:** 50 epochs, batch size of 32, with a validation split of 0.2.

**5. Model Evaluation**

* **Validation Accuracy:** Evaluated on a separate validation set to ensure model performance.
* **Test Accuracy:** Predicted outcomes on the test set and calculated accuracy.

**6. Issues Encountered**

**6.1 Column Alignment Errors**

* **Issue:** AttributeError related to column alignment between X\_train and X\_test.
* **Solution:** Converted X\_test to a DataFrame and used .reindex to align columns with X\_train.

**6.2 Handling NaN Values in Test Data**

* **Issue:** Test data contained NaN values.
* **Solution:** Dropped rows with NaN values and filled missing values appropriately.

**7. Final Steps and Results**

1. **Data Preprocessing:** Ensured test data had the same features as training data and handled missing values.
2. **Model Training:** Successfully trained the model and saved it.
3. **Model Evaluation:** Evaluated the model's accuracy on the test set.

**8. Conclusion**

The machine learning model for predicting asteroid hazard potential was successfully developed and evaluated. The project involved extensive data preprocessing to handle missing values and feature selection, followed by training and evaluating a neural network model. The final model demonstrated the ability to predict hazardous asteroids with satisfactory accuracy.