



NASA NEAREST OBJECT PREDICTION



CONTENTS

- Agenda
- Data Description
- Process
- Model 1, Model 2
- Result

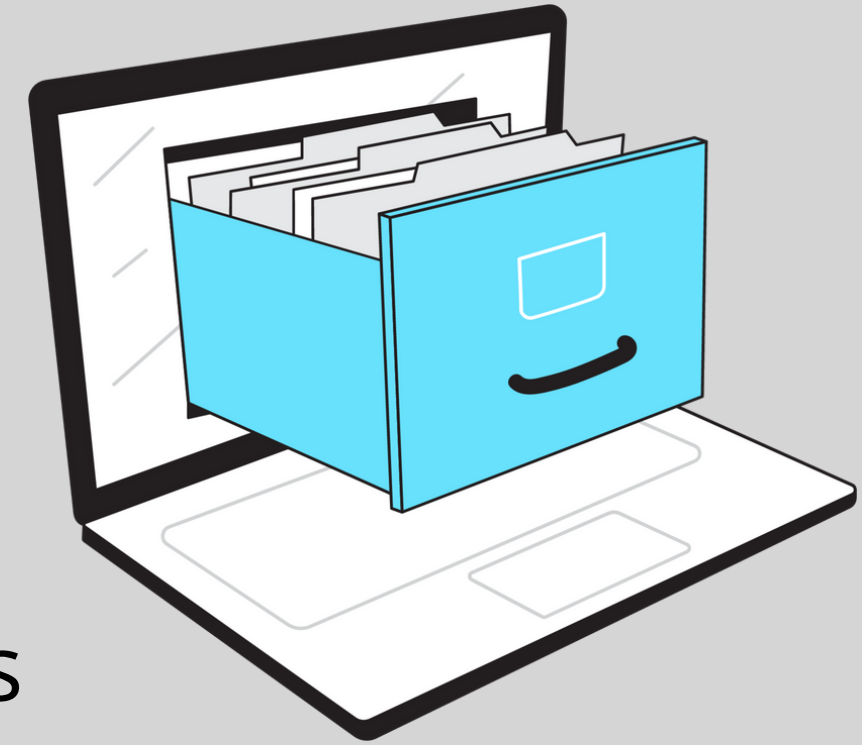


AGENDA

- There is an infinite number of objects in the outer space.
- Some of them are closer than we think. Even though we might think that a distance of 70,000 Km can not potentially harm us, but at an astronomical scale, this is a very small distance and can disrupt many natural phenomena. These asteroids can thus prove to be harmful.
- Hence, it is wise to know what is surrounding us and what can harm us amongst those. Thus, this dataset compiles the list of NASA certified asteroids that are classified as the nearest earth object.

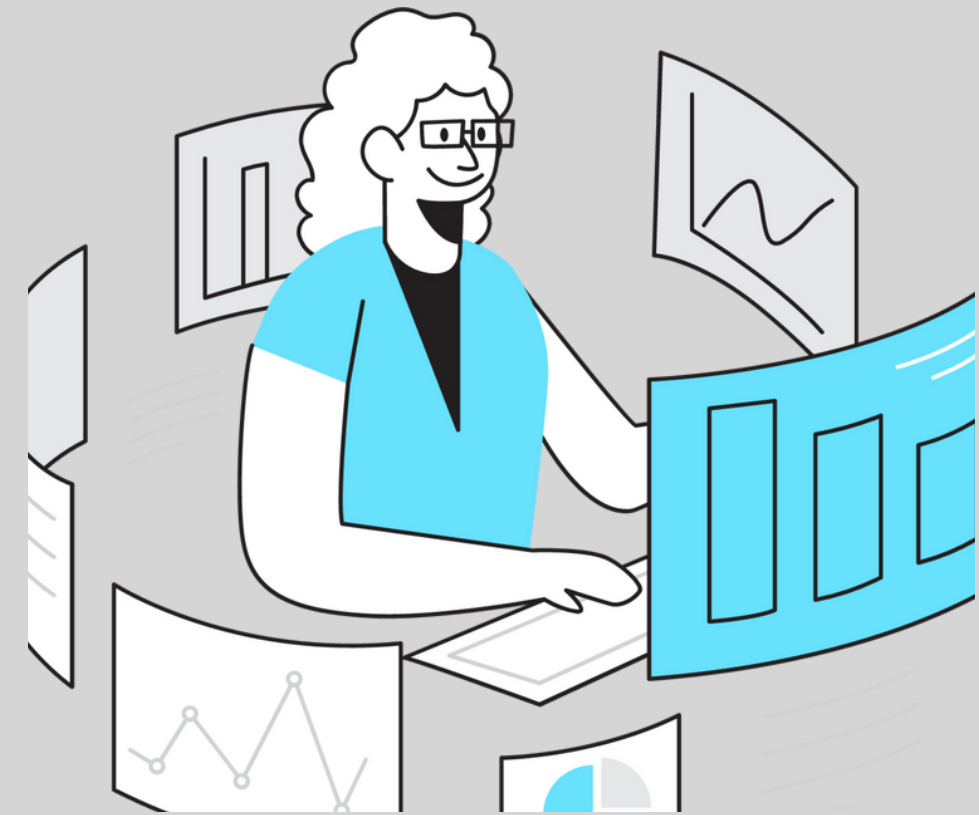
DATA DESCRIPTION

- **id** : Unique Identifier for each Asteroid
- **name** : Name given by NASA
- **est_diameter_min** : Minimum Estimated Diameter in Kilometres
- **est_diameter_max** : Maximum Estimated Diameter in Kilometres
- **relative_velocity** : Velocity Relative to Earth in Kmph
- **miss_distance** : Distance in Kilometres missed
- **orbiting_body** : Planet that the asteroid orbits
- **sentry_object** : Included in sentry - automated collision monitoring system
- **absolute_magnitude** : Describes intrinsic luminosity



Process

- Load dataset
- Inspecting the dataset
- Preprocessing the data
- Data Visualization
- Splitting the dataset into train and test sets
- Fitting and Making predictions RandomForest Model
- Fitting and Making predictions DecisionTree Model
- Making predictions and evaluating performance



LOAD DATASET

It is the data that we need to load for starting any of the ML project. With respect to data, the most common format of data for ML projects is CSV (comma-separated values).

INSPECTING THE DATASET

To ensure that you're dealing with the right information you need a clear view of your data at every stage of the transformation process

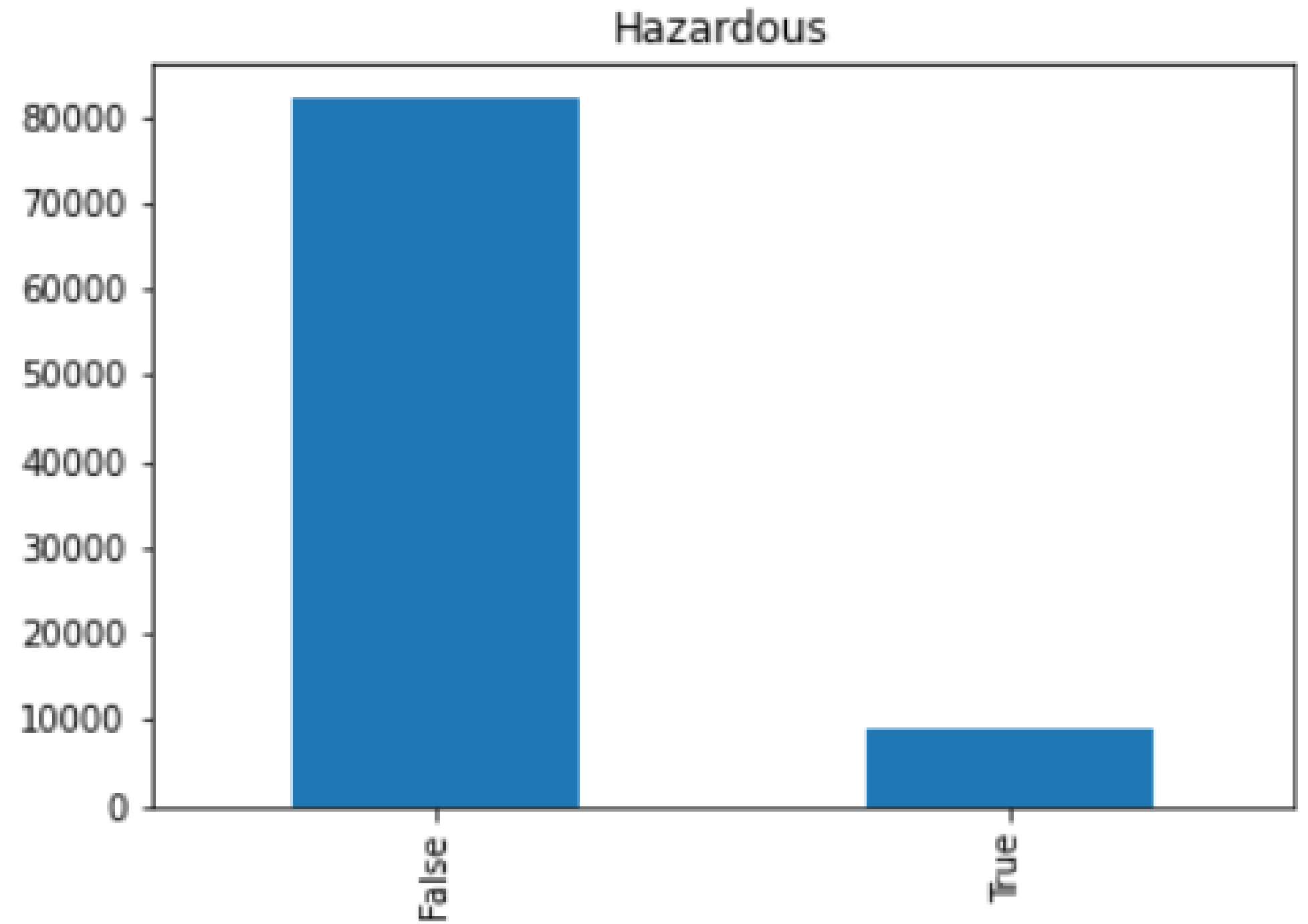
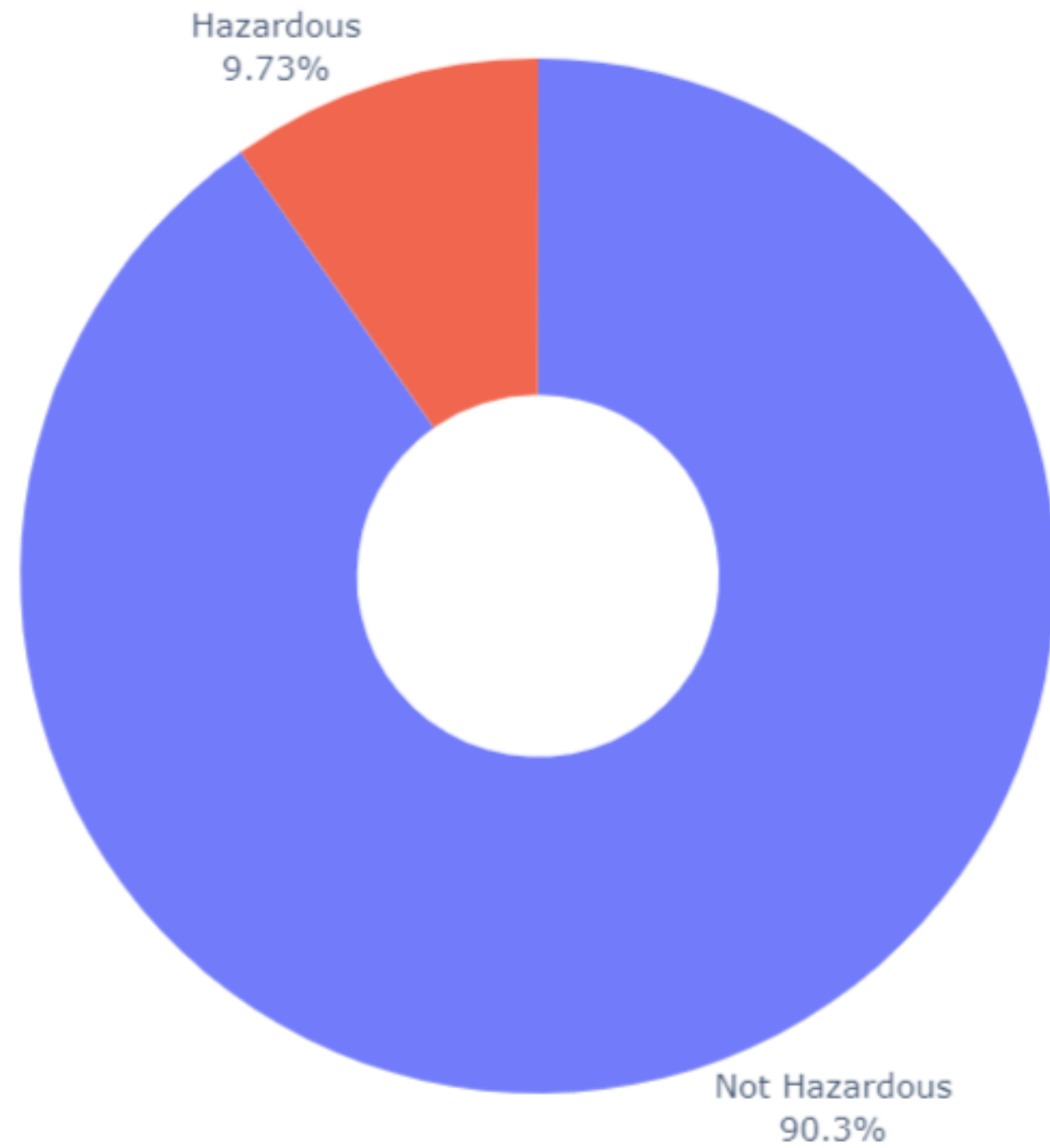
PREPROCESSING THE DATA

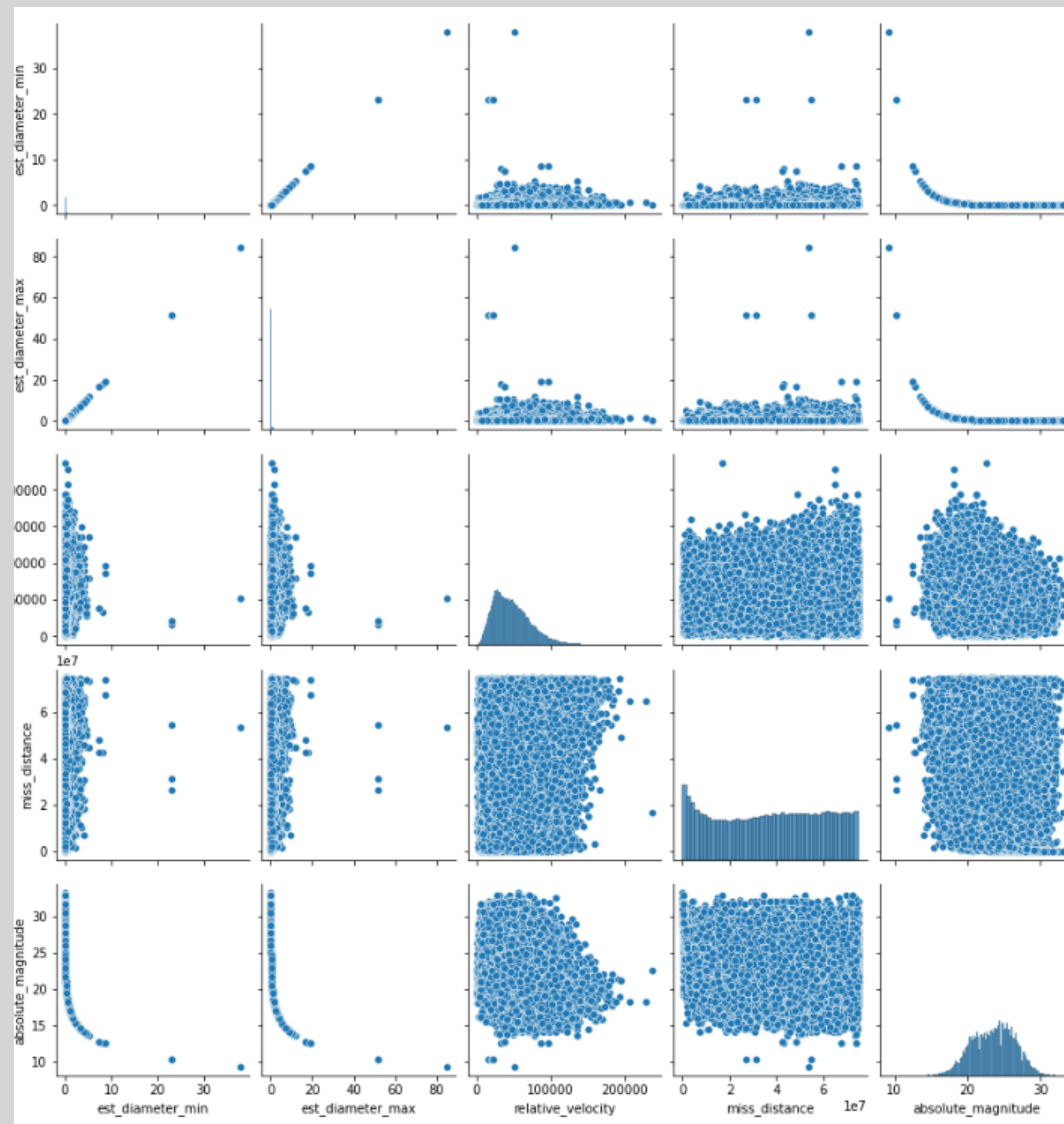
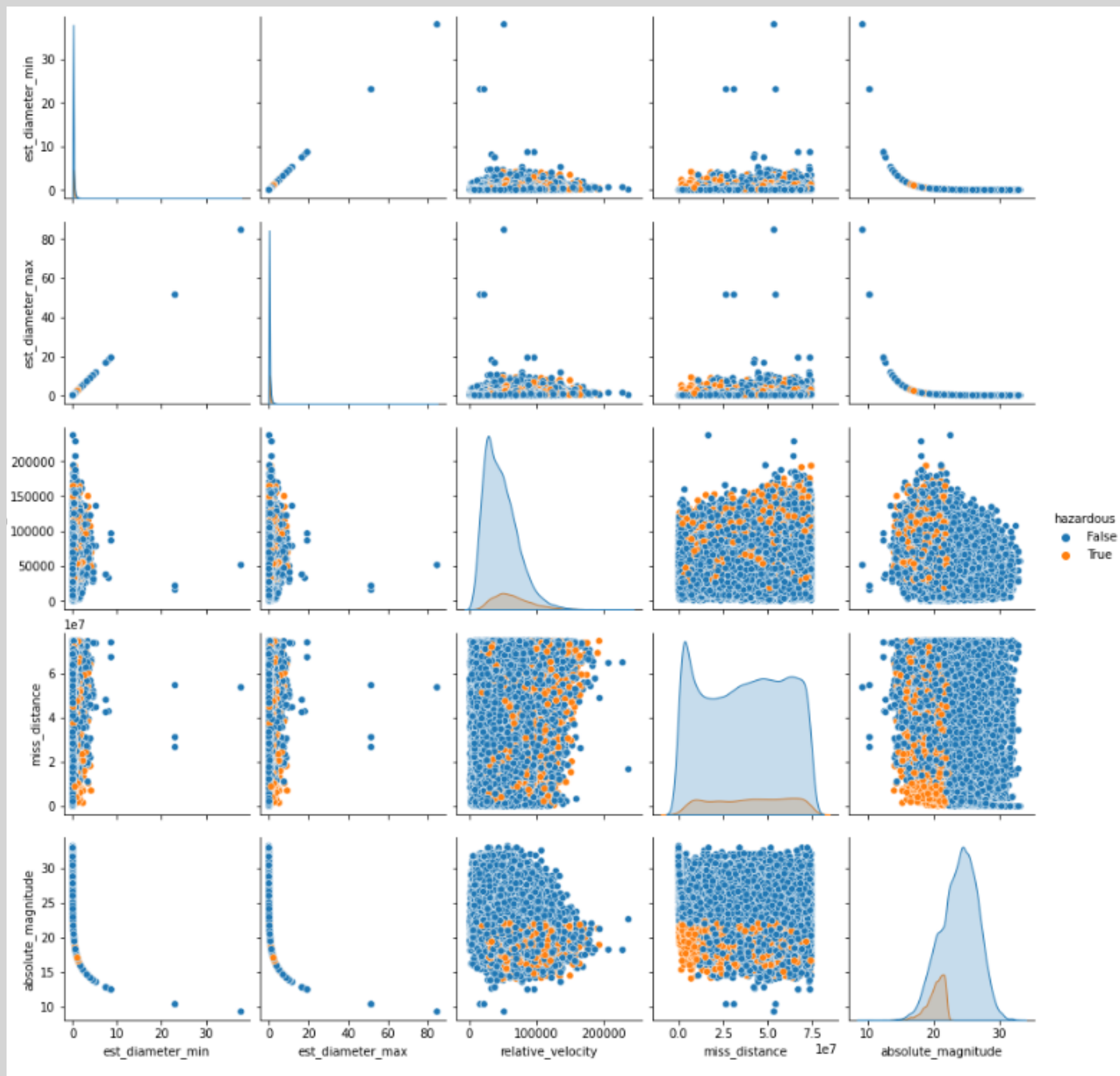
Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model

SPLITTING THE DATASET INTO TRAIN AND TEST SET

Separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

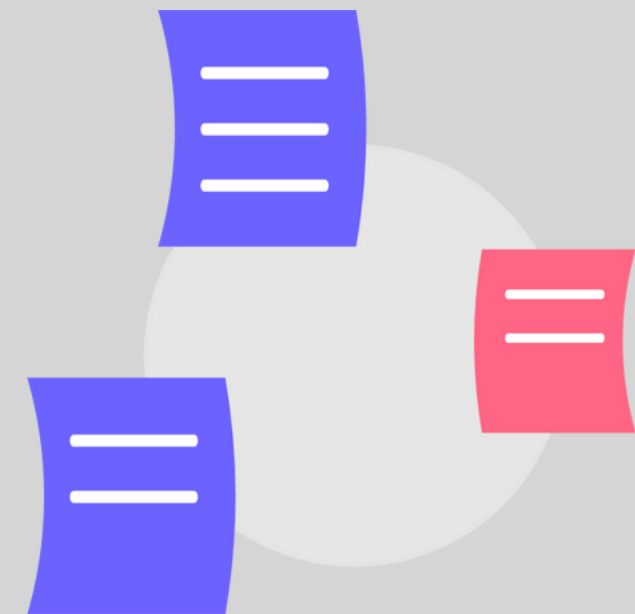
DATA VISUALIZATION





FITTING AND MAKING PREDICTION USING RANDOM FOREST MODEL

- A random forest classifier. A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.
- Fitted random forest classifier
- Predict the model using "X test"
- Calculate the Accuracy score, Confusion matrix, Classification report
- Calculated for both the Train and Test dataset



FITTING AND MAKING PREDICTION USING DECISION TREE CLASSIFIER

- Decision Tree is a Supervised Machine Learning Algorithm that uses a set of rules to make decisions, similarly to how humans make decisions.
- Fitted Decision Tree classifier
- Predict the model using "X test"
- Calculate the Accuracy score, Confusion matrix, Classification report
- Calculated for both the Train and Test dataset



RESULT

- Both the models performed well on the testing data.
- Concatenate the stored training and testing data in a separate dictionaries for the comparision of algorithms
- The concatenation is done by using the pandas library
- Finally displays the result of all the algorithms after concatenation
- By observing all the result best model can be choosed
- Random forest classifier is the best classification model for the dataset uploaded
- It has the accuracy score of 0.89 for the test sets



