HAZARDOUS ASTEROID PREDICTION USING ML

Team Achievers

I.Yasas Chandra - 19BCE7075

G.Sreeharsha - 19BCD7114

P.P.K.Satya Raj - 19BCD7041

J.Bhargav - 19BCE7460

Asteroids History - Introduction

Asteroids are leftovers from the formation of our solar system about 4.6 billion years ago. Early on, the birth of Jupiter prevented any planetary bodies from forming in the gap between Mars and Jupiter, causing the small objects that were there to collide with each other and fragment into the asteroids seen today.

Asteroids Killed Dinosaurs

Sixty-six million years ago, a mountain-size asteroid slammed into Earth just off the coast of Mexico's Yucatán Peninsula, dooming the dinosaurs and leading to their extinction.

The collision was cataclysmic, triggering tsunamis that swamped vast swaths of coastline and firestorms that may have raged across the entire globe(opens in new tab). The impact also blasted huge amounts of dust and vaporized rock into the air, which, along with the soot from all those fires, blocked the sun for long stretches.

To make matters worse, much of that vaporized rock was rich in sulfur, generating sulfuric acid aerosols that rained back down on Earth and acidified the oceans.

Abstract

In this Project, we applied different machine learning methods namely, Naive bayes, Logistic Regression, Random Forest, decision tree, k- nearest neighbors, Support Vector Machine(SVM) and Deep neural networks – for identifying potential hazardous and non-hazardous asteroids and features responsible for claiming an asteroid to be hazardous.

So far as, asteroids are concerned, there are many asteroids called near-earth asteroids, but all are not hazardous. So, our target in this paper is to identify those hazardous asteroids and classify them with non-hazardous types. For this, we choose many machine learning models By these results obtained, Now we are going to create a mobile app using flutter for predicting an asteroid is hazardous or not. And name of the app is "Asteroid".

Objectives

A focus should be placed on compiling and training large datasets of asteroid observations for improving the accuracy of these machine learning tools already at our disposal. It has been seen that now-a-days, machine learning is the one of the most important technique for predicting or classifying the dataset. So here, I use some of the machine learning model and later compare those results to show.

Challenges in existing system

- Due to the presence of vast and large dataset, there was a need to eliminate unnecessary columns. Also there is a large Prescence of missing values and outliers in existing necessary columns. Also a model need to be developed with large accuracy keeping preceding points in mind with least loss.
- A model is said to be a good machine learning model if it generalizes any new input data from the problem domain in a proper way. This helps us to make predictions in the future data, that the data model has never seen. Now, suppose we want to check how well our machine learning model learns and generalizes to the new data. For that, we have overfitting and underfitting, which are majorly responsible for the poor performances of the machine learning algorithms
- When a app is developed to predict potential hazardous asteroids there is a need to gather large information about asteroid. User may not have such information about asteroid. Precisely, there is a need to use API to generalize information about asteroid.

Proposed Methodology

- **Data collection:**

This data has been collected from the official website of Jet Propulsion Laboratory of California Institute of Technology which is an organization under NASA, and maintained by them. Dataset Contains 4688 rows and 40 features.

- **Data Visualization:**

We need to delete redundant features correlated with each other and after checking of duplicate

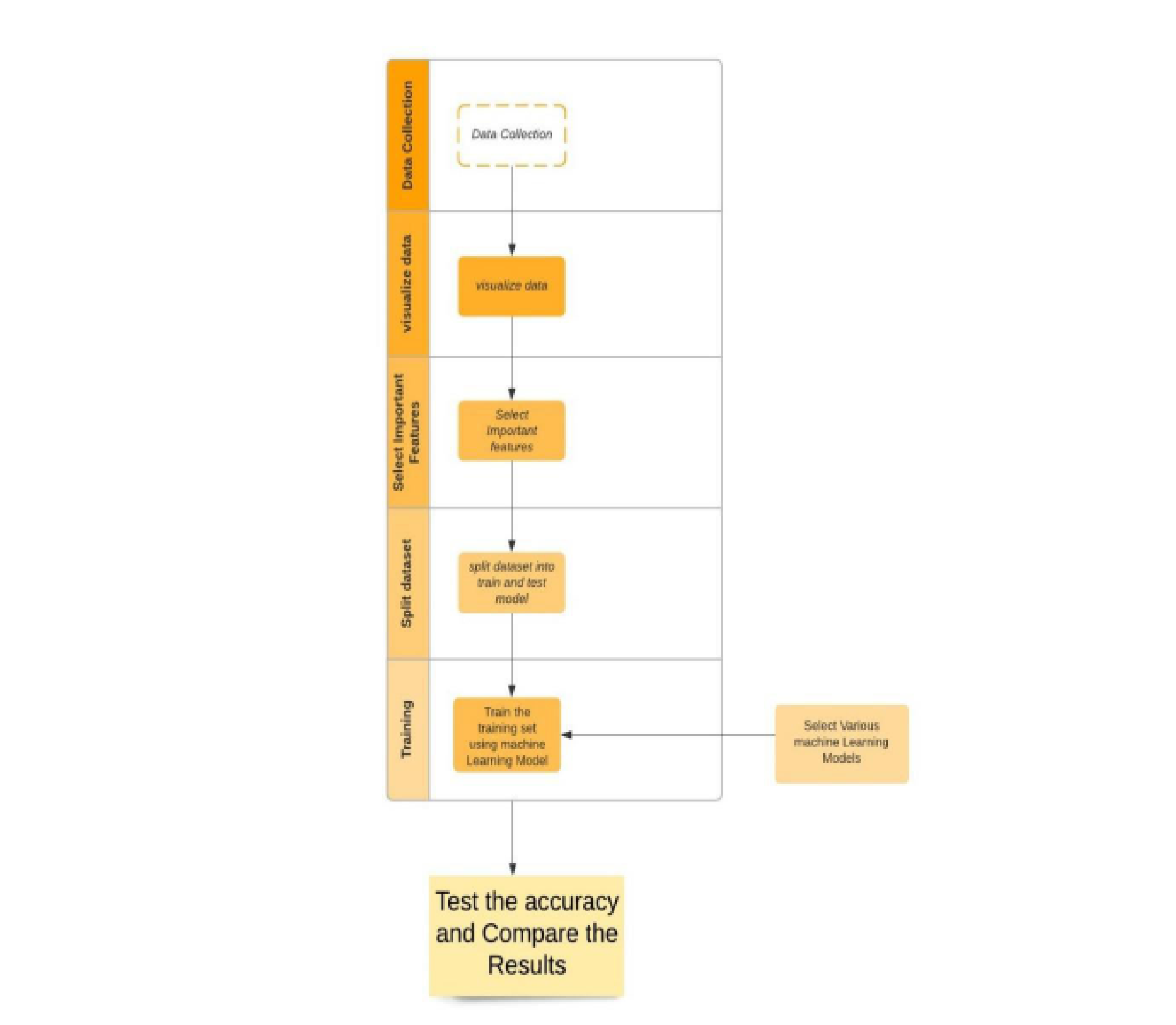
values and missing values, we need to proceed for Features Selection.

- Features selection:

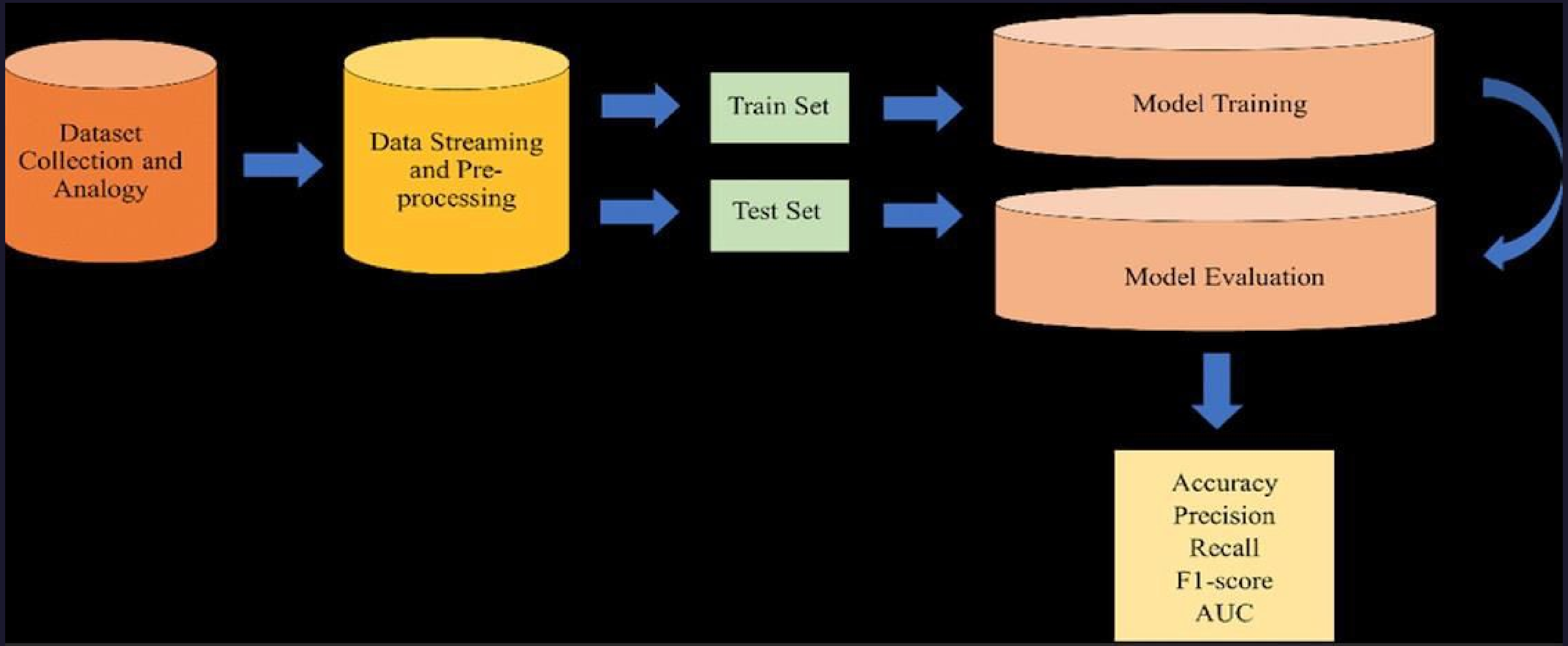
Some insignificant features need to be identified and removed before splitting training and testing dataset.

- Split dataset:

Then, the dataset should be divided into test and train set. We are precisely identifying baseline accuracy and trying to create split dataset as dataset is imbalanced.



Architecture Diagram



Machine Learning Models

- We intend to use Supervised Machine Learning models to classify Hazardous Asteroids and Non-Hazardous Asteroids.
- The following models are used:
- **Naïve Bayes** is one of the fast and easy ML algorithms to predict a class of datasets. It can be used for Binary as well as Multi- class Classifications.
- **Logistic Regression** is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, any more complex extensions exist
- **Random Forest** The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method.

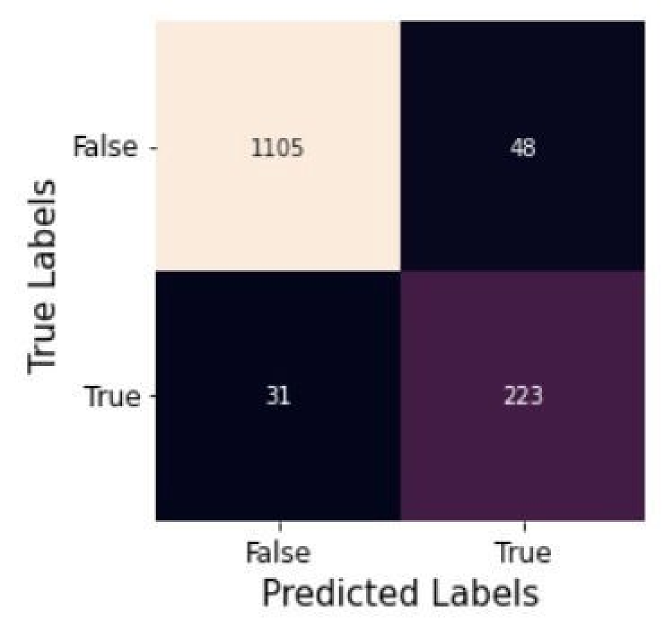
- **Decision tree** A decision tree combines some decisions. a decision tree is fast and operates easily on large data sets, especially the linear one.
- **k-Nearest Neighbors (** KNN) is Classification Algorithm which classifies the data point on how its neighbor is classified.
- **Deep neural network** (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. There are different types of neural networks but they always consist of the same components: neurons, synapses, weights, biases, and functions
- **Support Vector Machine**(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points

Implementation Details

- **Naïve Bayes:**

```
Accuracy of train: 0.949390243902439
Accuracy of test: 0.9438521677327647
```

	precision	recall	f1-score	support
0	0.97	0.96	0.97	1153
1	0.82	0.88	0.85	254
accuracy			0.94	1407
macro avg	0.90	0.92	0.91	1407
weighted avg	0.95	0.94	0.94	1407

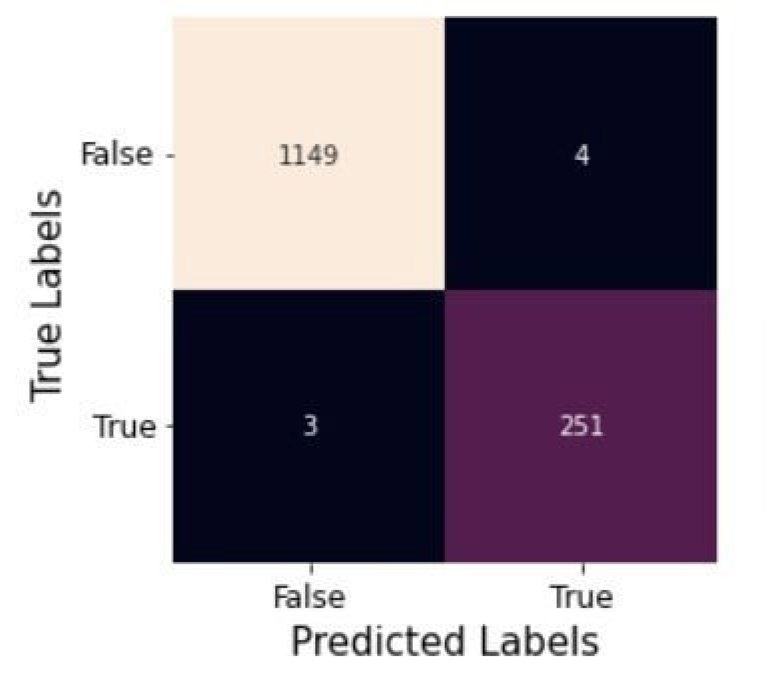



```
Precision = 0.8977947481939608
Recall = 0.9181611134254359
Accuracy = 0.9438521677327647
F1 Score = 0.9075054608999562
```

Random Forest

```
Accuracy of train: 1.0
Accuracy of test: 0.9950248756218906
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1153
1	0.98	0.99	0.99	254
accuracy			1.00	1407
macro avg	0.99	0.99	0.99	1407
weighted avg	1.00	1.00	1.00	1407

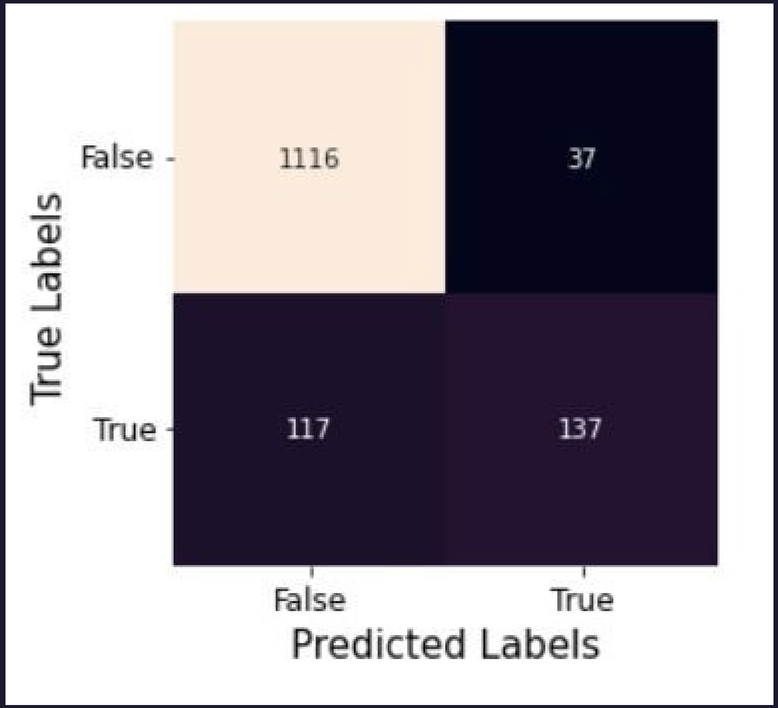



```
Precision = 0.9908547794117647
Recall = 0.9923598828116997
Accuracy = 0.9950248756218906
F1 Score = 0.991605333924287
```

KNN

Accuracy of train: 0.913109756097561
Accuracy of test: 0.8905472636815921

	precision	recall	f1-score	support
0	0.91	0.97	0.94	1153
1	0.79	0.54	0.64	254
accuracy			0.89	1407
macro avg	0.85	0.75	0.79	1407
weighted avg	0.88	0.89	0.88	1407

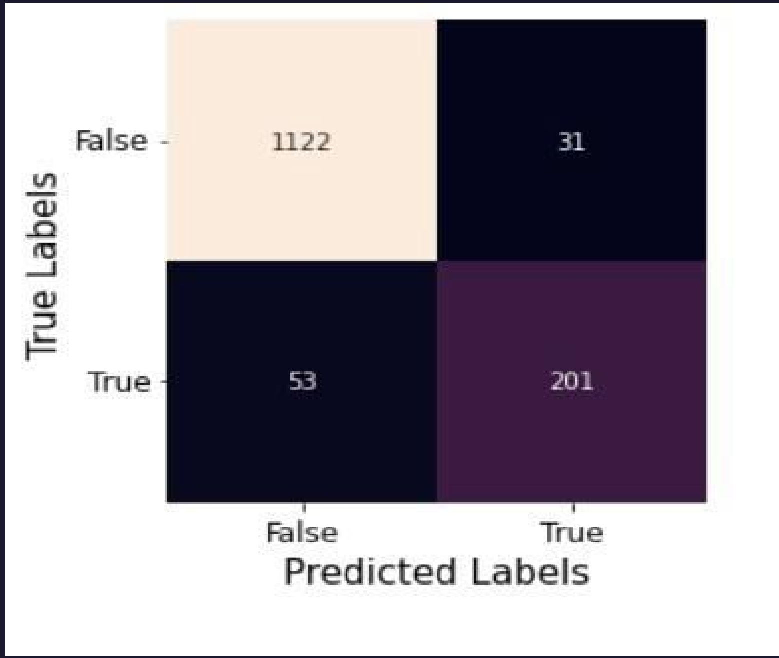


Precision = 0.8462329054450877
Recall = 0.7536399396302695
Accuracy = 0.8905472636815921
F1 Score = 0.7878218737025171

SVM

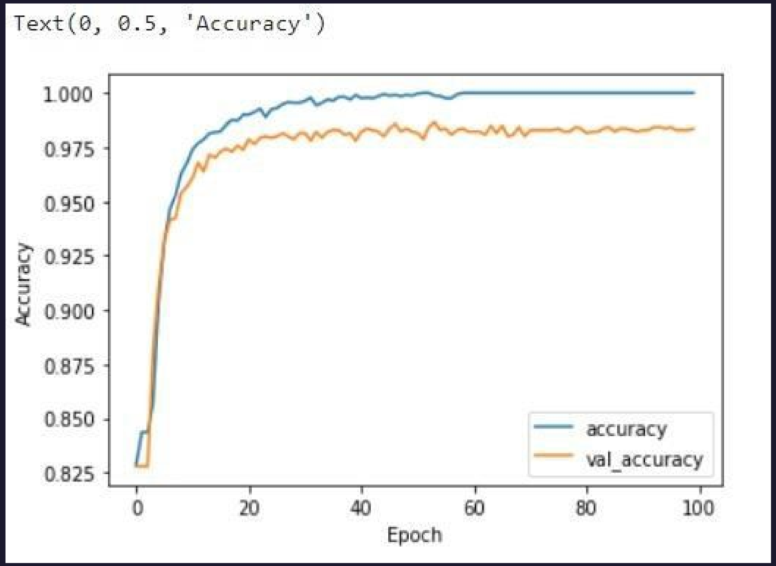
```
Accuracy of train: 0.961890243902439
Accuracy of test: 0.9402985074626866
```

	precision	recall	f1-score	support
0	0.95	0.97	0.96	1153
1	0.87	0.79	0.83	254
accuracy			0.94	1407
macro avg	0.91	0.88	0.90	1407
weighted avg	0.94	0.94	0.94	1407




```
Precision = 0.9106364636830522
Recall = 0.8822260996646885
Accuracy = 0.9402985074626866
F1 Score = 0.8955390098001782
```

Deep Neural Network

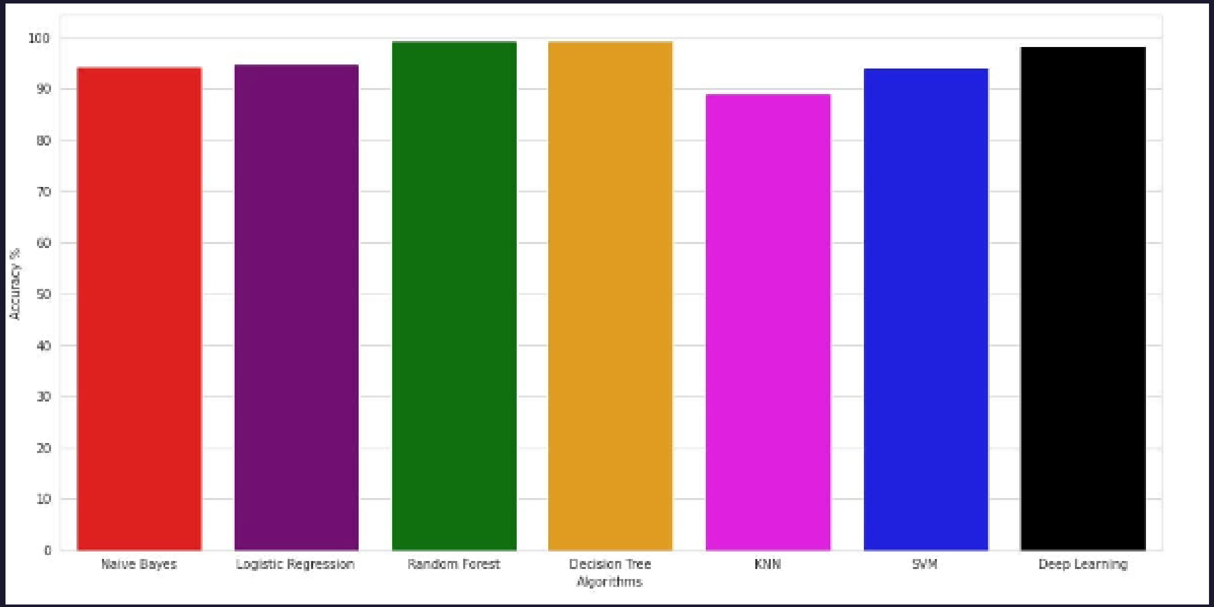


	precision	recall	f1-score	support
0	0.99	0.99	0.99	1165
1	0.96	0.95	0.95	242
accuracy			0.98	1407
macro avg	0.97	0.97	0.97	1407
weighted avg	0.98	0.98	0.98	1407

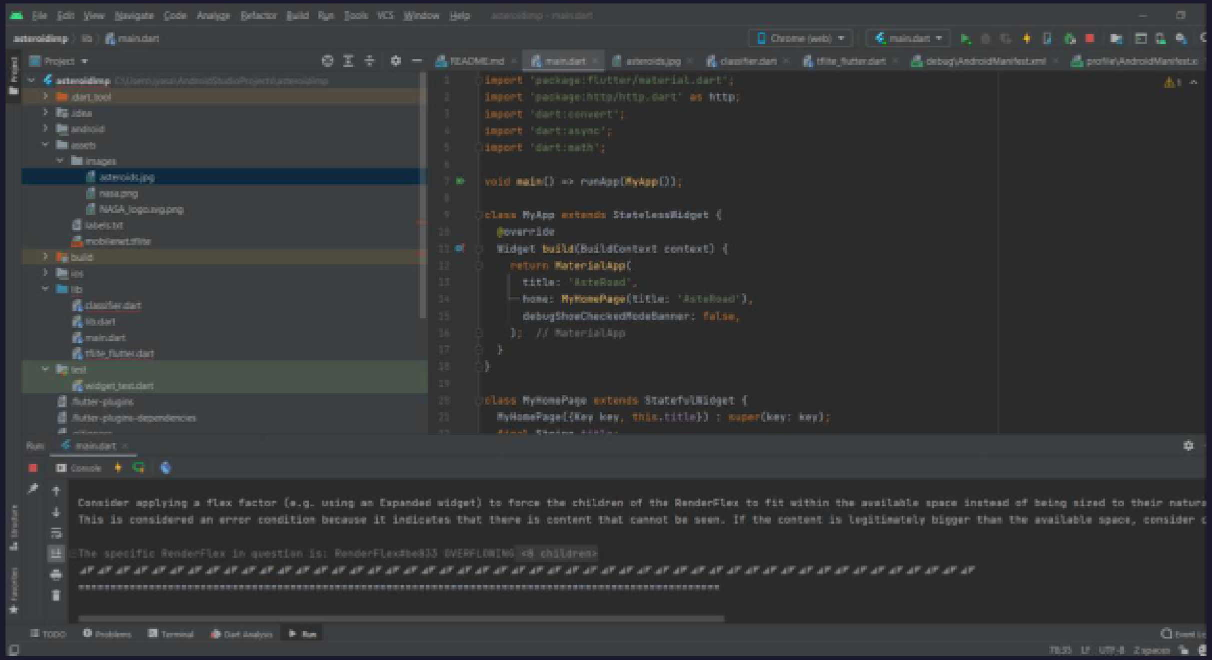
Precision = 0.973514429414799
Recall = 0.9688486503742064
Accuracy = 0.9836531627576404
F1 Score = 0.9711622004806746

Final Comparison

```
Naive_Bayes_accuracy            0.943852
Logistic_Regression_accuracy    0.948117
Random_Forest_accuracy          0.995025
Decision_Tree_accuracy          0.993603
KNN_accuracy                    0.890547
SVM_accuracy                    0.940299
Deep_Learning_accuracy          0.983653
dtype: float64
```



Flutter App




AsteRoad
NASA
Welcome to AsteRoad. Get some useful information about asteroids nearby/approaching earth.
Asteroid name: 433 Eros (A898 PA)
Diameter: 48.7573371375 km
Hazardous?: false
Closest Approach Date: 1900-12-27

Conclusion

- In this project, we have done this classification of hazardous and non- hazardous through various machine learning techniques. In my opinion, this paper will help to identify the newly discovered near earth asteroids if that is hazardous or not. For larger objects, which impact much less frequently but would do far more damage, it is fair to expect we would receive some warning. From the above after performing all the different machine learning models the results indicate that the Random Forest technique gives the best results in terms of prediction followed by Decision Tree as there is only 0.002 accuracy difference between Decision Tree and Random Forest.

Can We prevent Asteroids?

- Deflection for asteroid

If, or really when, an asteroid is found to be on a collision course with Earth, what do we do? This is not a topic you want to first consider when an asteroid is bearing down on Earth with the capability of wiping out a city, or creating a gigantic tsunami. There are a variety of possible asteroid deflection techniques in various states of readiness, but all need more development and testing. There are fewer options with short warning times and/or larger objects, and more options with additional warning time and smaller objects.

Project Prepared under the guidance of:

Dr.Monali Bordoloi

Assistant Professor Sr Grade 1