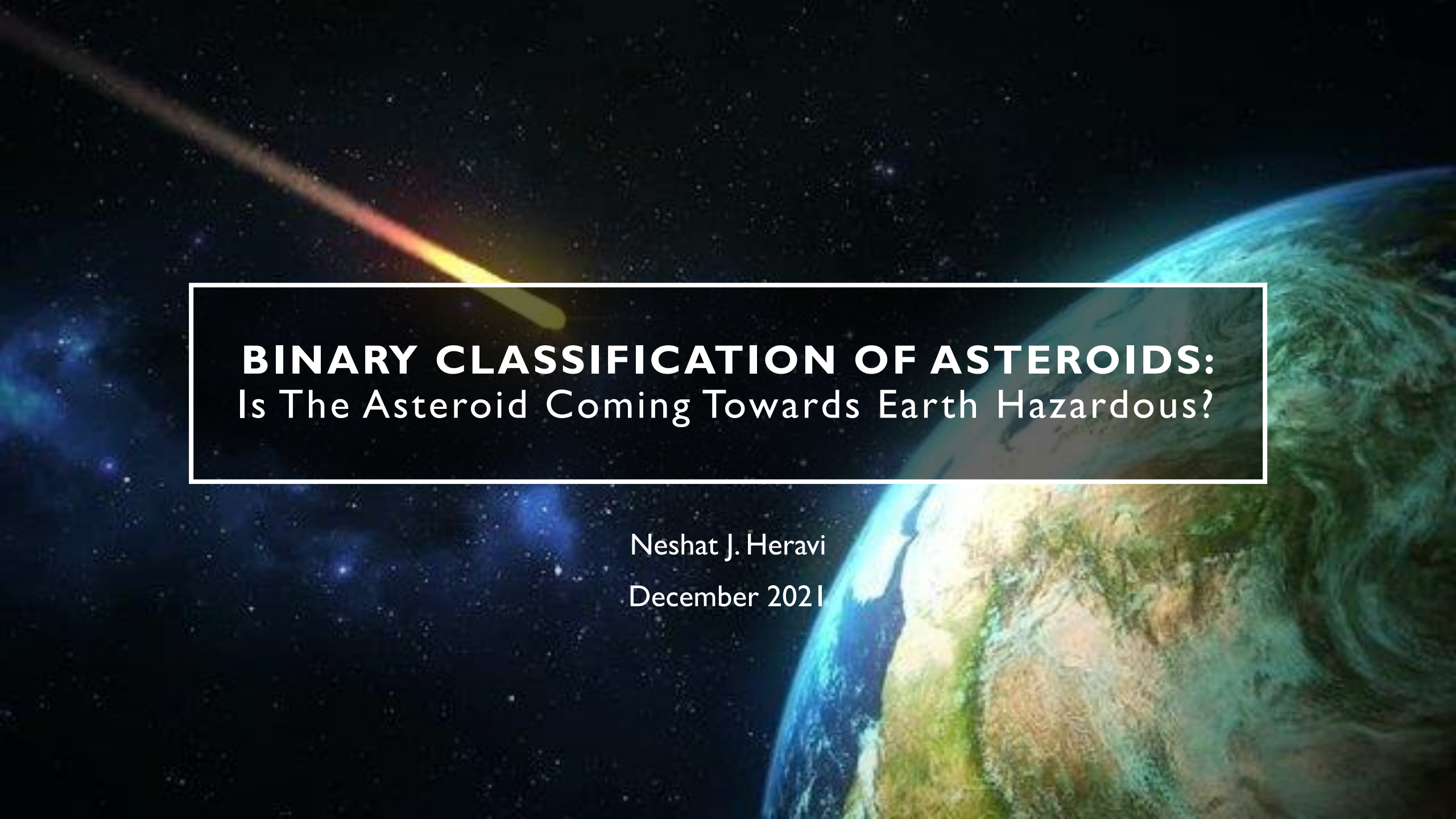


BINARY CLASSIFICATION OF ASTEROIDS:
Is The Asteroid Coming Towards Earth Hazardous?

Neshat J. Heravi

December 2021

REMEMBER DINOSAURS?

Introduction

- There are massive numbers of asteroids and comets orbiting the Sun. A tiny fraction of them can follow a path which passes by Earth. These Near-Earth Objects (NEOs) which have high traveling speeds can be the size of a pebble or a mountain!

Objective

- The objective is to build a reliable classification model to be able to identify if an asteroid is hazardous or not!

Goal

- NASA and government agencies in charge can use this information as part of their planetary-defense strategies. Being able to predict if an asteroid is hazardous or not can save cost and resources. Besides, our lives may depend on it!

RECENT HEADLINES

Potentially hazardous asteroid to fly by Earth this week

The International Astronomical Union lists over 1,500 potential hazardous asteroids.

Football field-sized asteroid to approach Earth on Monday night, one of several in the coming weeks

NASA's 'Eyes on Asteroids' Reveals Our Near-Earth Object Neighborhood

Nasa asteroid warning 2021: Eiffel Tower-sized asteroid heading towards earth in December - should we worry?

NEWS

Massive, 'potentially hazardous' asteroid to pass by Earth Saturday, December 10

DATA SCIENTIST SOLUTION

Use Machine Learning to save the Earth!

Using data from NASA open API we can build a predictive model!

A binary classification model can predict if the approaching asteroid is hazardous or not!

METHODOLOGY

Data Cleaning & Preprocessing

Asteroids data from NASA open API on Kaggle

Python Tools:
Numpy, Pandas, Scikit-learn,
Matplotlib, Seaborn

Feature Engineering & Building Models

Handling Class Imbalance:
SMOTE (over-sampling technique)

Train-Test Split:
Data was split into train and test sets (80/20)

Models:
Logistic Regression, SVM, Decision Tree,
Random Forest, XGBoost

Model Tuning & Evaluation

Hyperparameter Tuning:
GridSearchCV

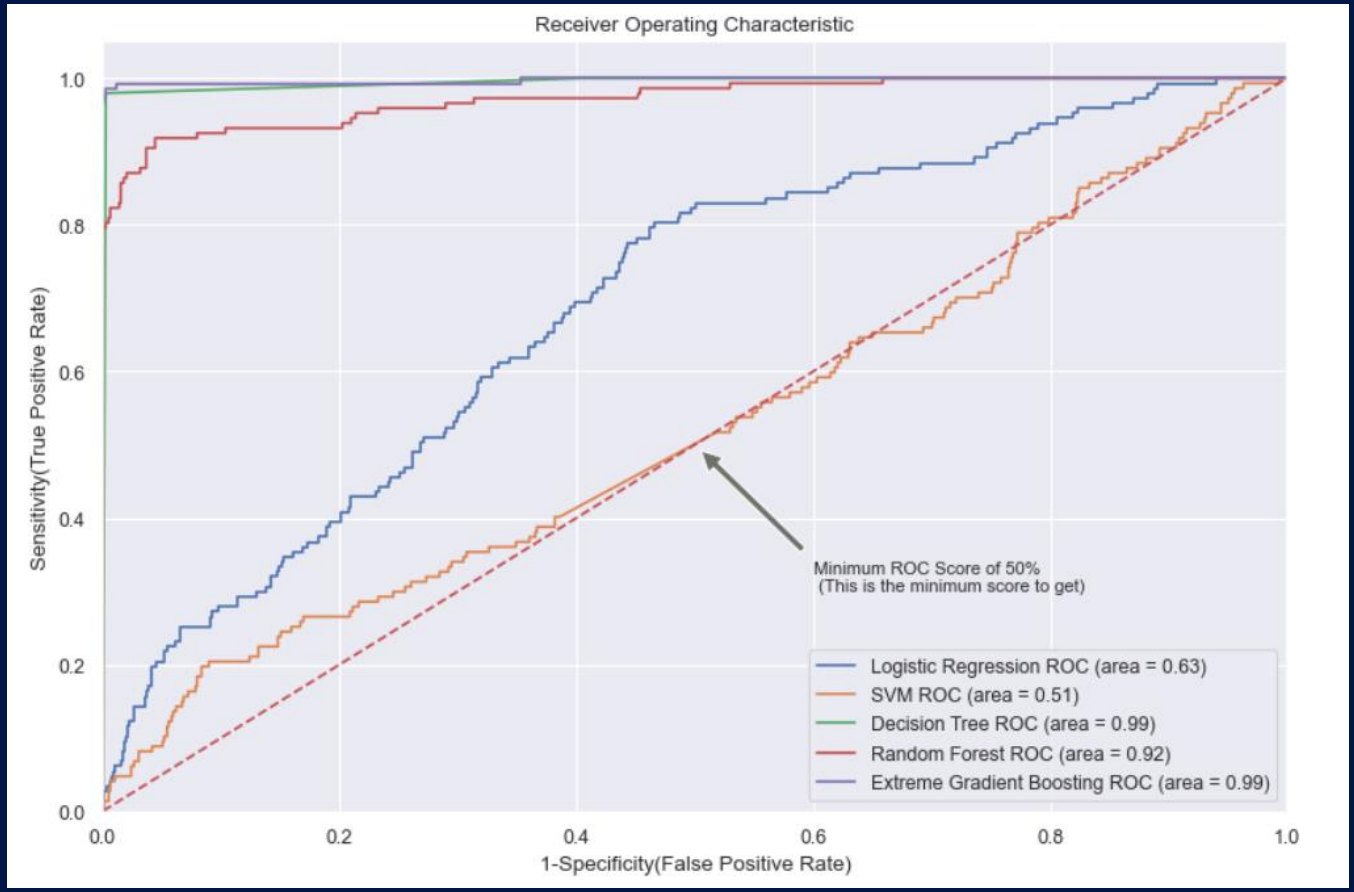
Evaluation Metric:
Accuracy, Recall, Precision, F1 Score, ROC-AUC curve and AUC score

Feature Importance:
From Decision Tree model

MODELS COMPARISON

Model Performance From Best to Worst:

- Decision Tree & XGBoost (tied)
- Random Forest
- Logistic Regression
- SVM



FINAL MODELS

	Decision Tree	XGBoost
Accuracy	99.47 %	99.47 %
Recall	97.96 %	98.64 %
Precision	98.63 %	97.97 %
F1 Score	0.98	0.98
AUC Score	0.99	0.99

Pros and Cons: Decision Tree Model is more interpretable. However, XGBoost Model is more stable and require less maintenance!

CONCLUSIONS

- Multiple ML **binary classification models** were built, trained, tuned and evaluated based on their performance metrics.

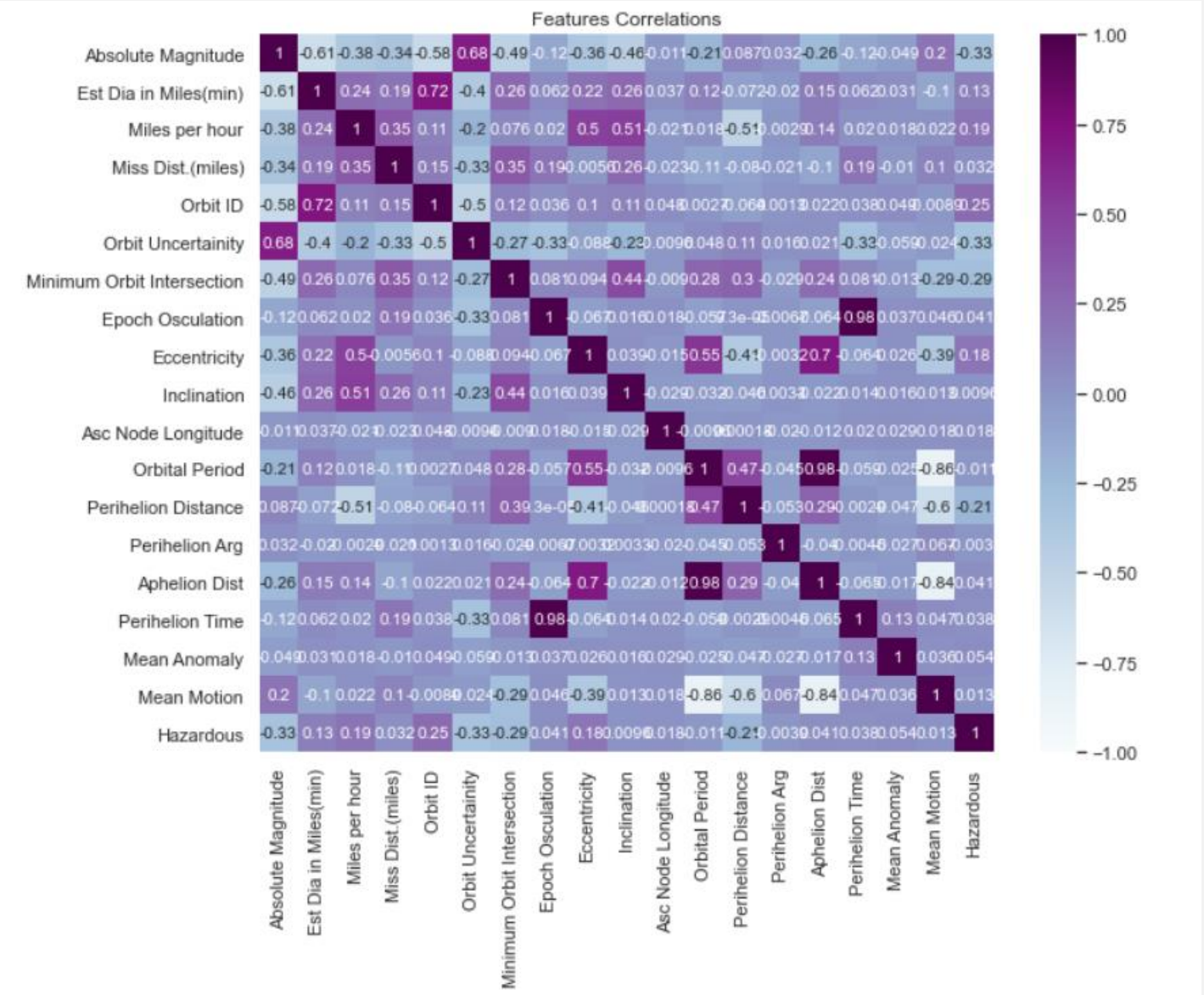
- Two best performed models were **Decision Tree** and **XGBoost** with **99.47%** accuracy, and **0.99** AUC score for both models.

- **Feature Importance** extracted from the Decision Tree model implied the importance of Asteroid's **Minimum Estimated Diameter**, and the **Minimum Orbit Intersection** features for predictions.


THANK YOU

APPENDIX:

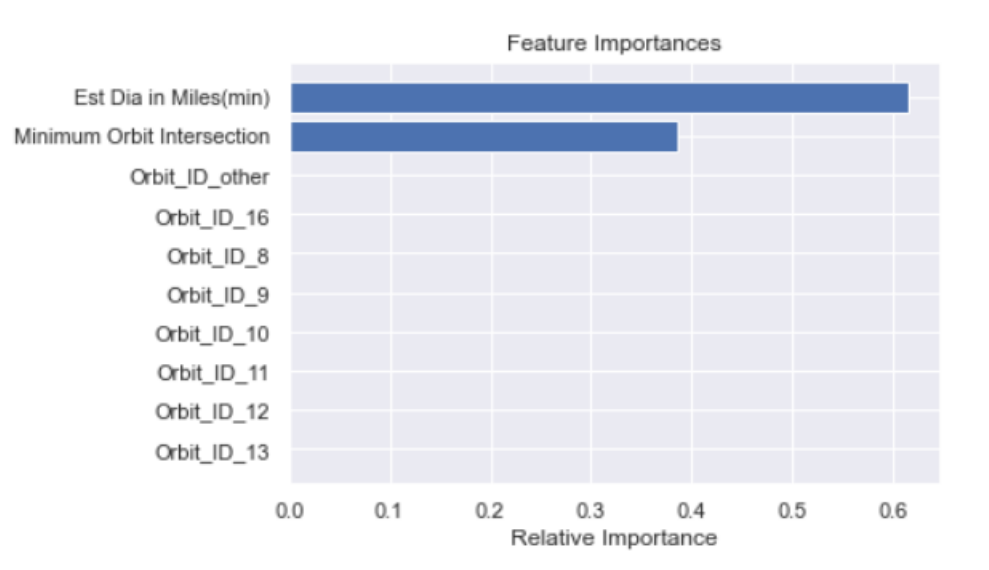
Feature Correlation Heatmap



APPENDIX:

Class Imbalance

Feature Importance from Decision Tree Model



APPENDIX:

Confusion Matrices for different models