

ASTEROID THREAT CLASSIFICATION

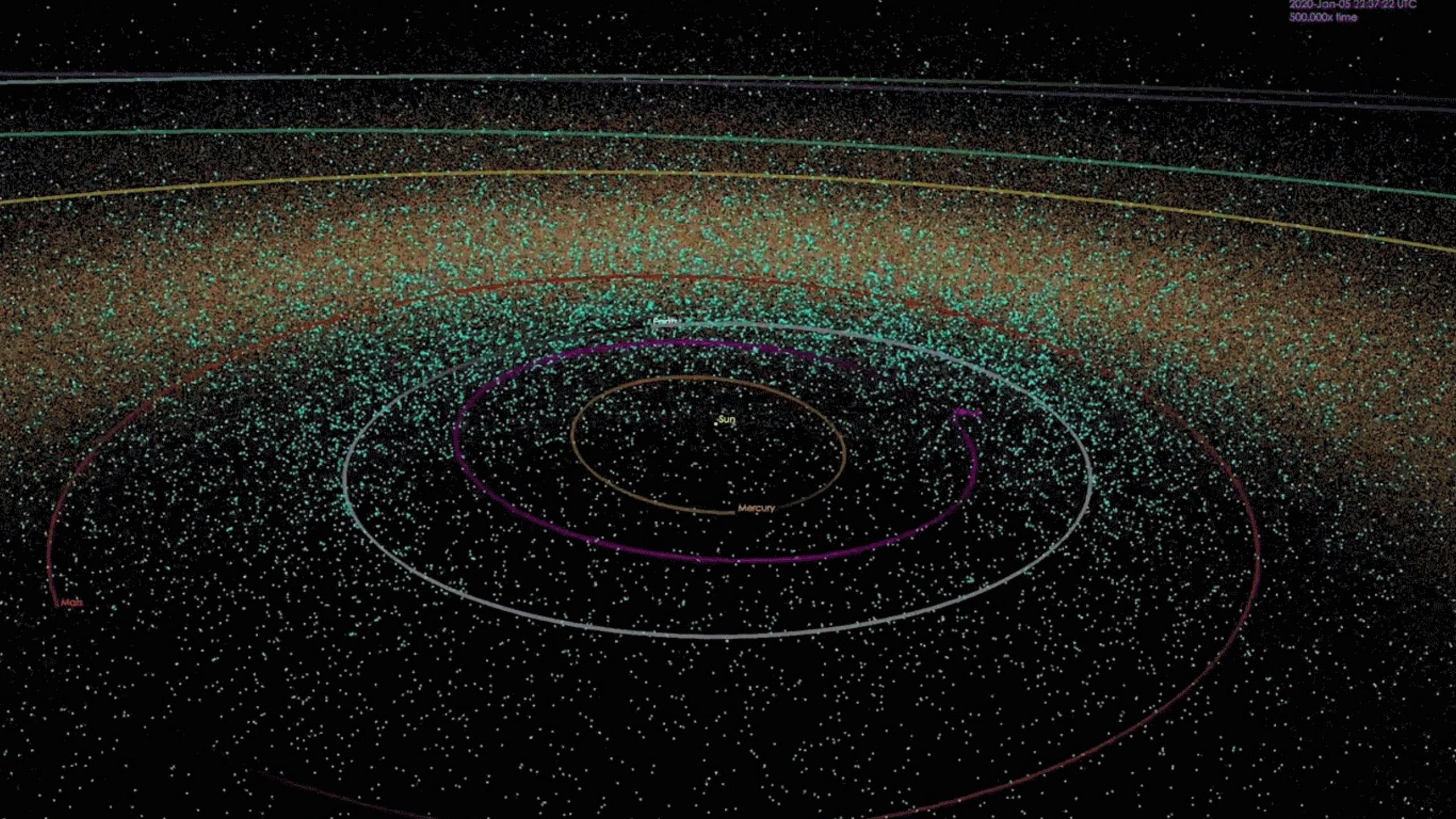
Adithya Viswanathan

MOTIVATION

Astronomy

Danger

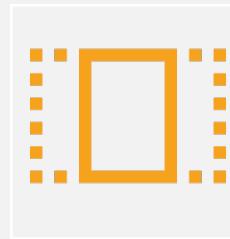
Classification



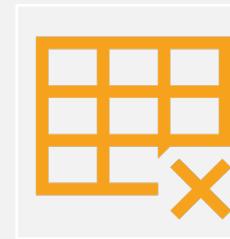
DATASET



Asteroid Threat



16 columns



Over 15000 rows

STEPS

1

Data
Preparation

2

Data Analysis

3

Data
Preprocessing

4

Model Fitting

5

Metrics

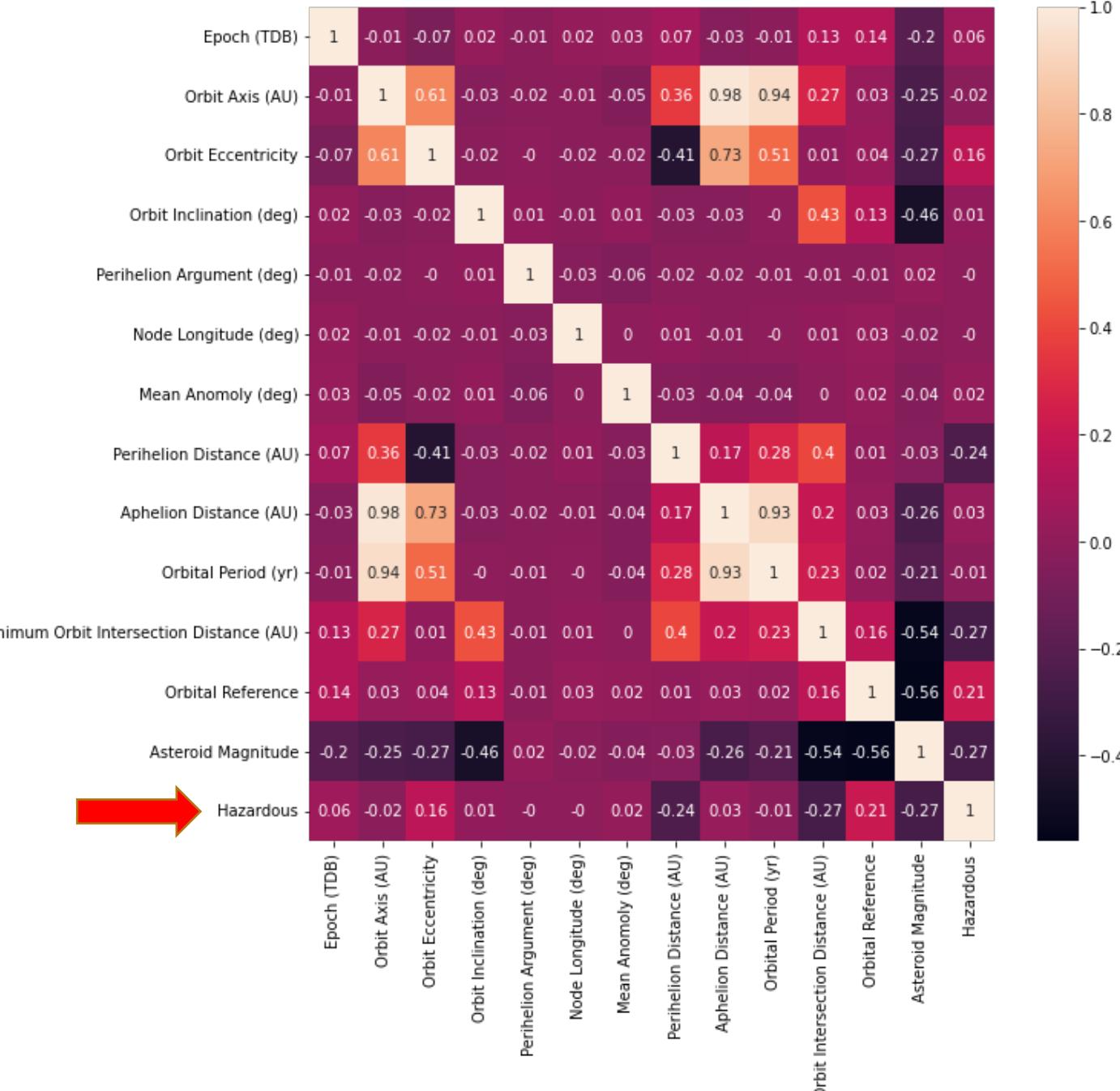
OBSERVATIONS

Highly imbalanced
dataset (11% true,
89% false)

Feature
Correlations?

Metrics?

Correlation matrix of Asteroid data



FEATURE SELECTION

- Minimum Orbit Intersection Distance (AU)*
- Asteroid Magnitude*
- Perihelion Distance (AU)
- Epoch
- Orbit Eccentricity
- Orbital Reference
- Asteroid Type

MODELS

K-Nearest
Neighbors

Decision
Tree

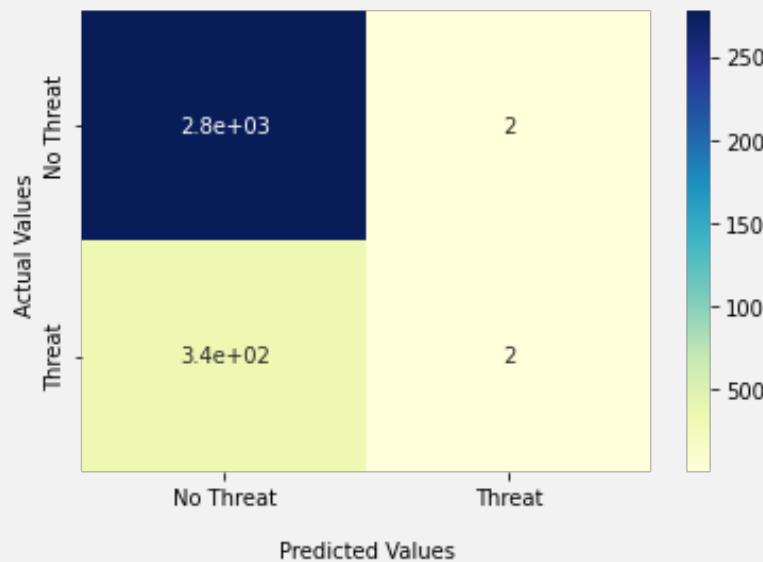
Naïve
Bayes

Logistic
Regression

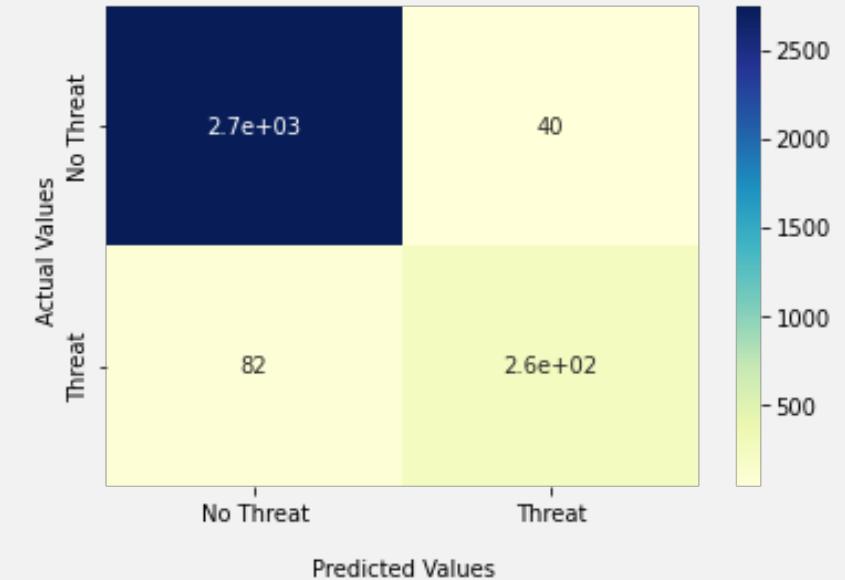
Neural
Network

RESULTS

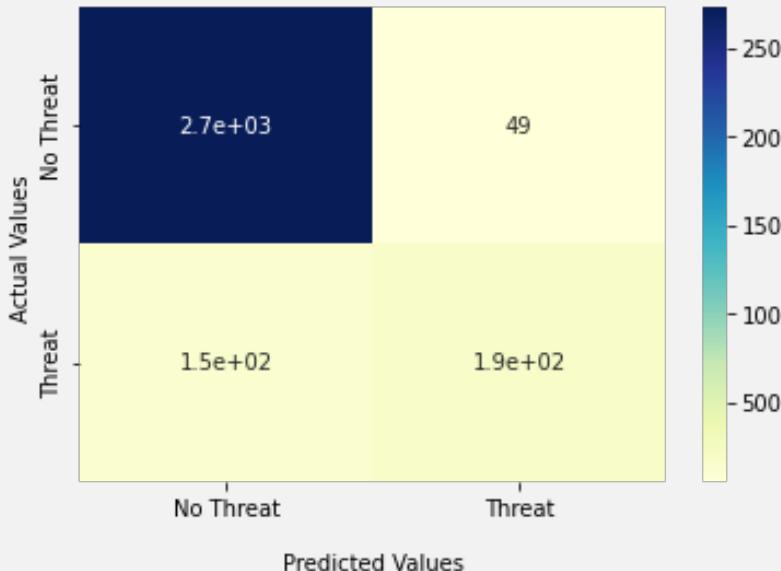
K-Nearest Neighbors Confusion Matrix



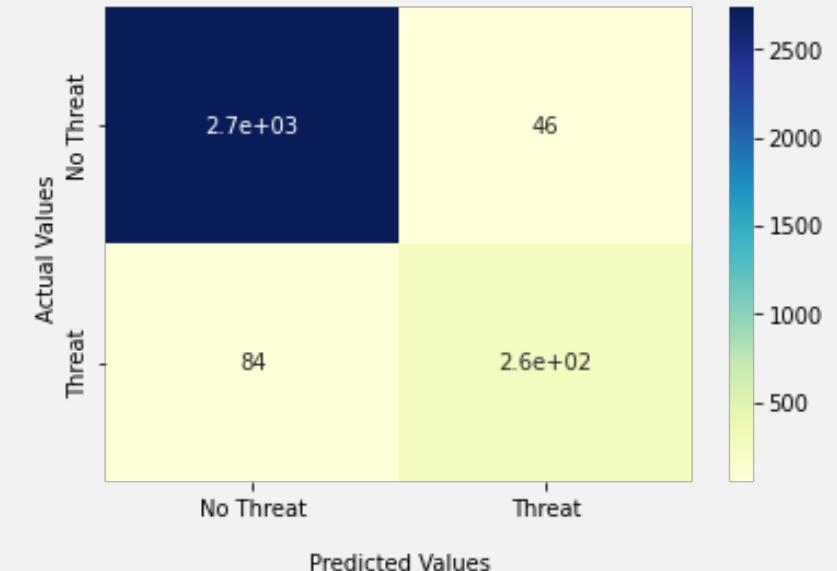
Naïve Bayes Confusion Matrix



Logistic Regression Confusion Matrix

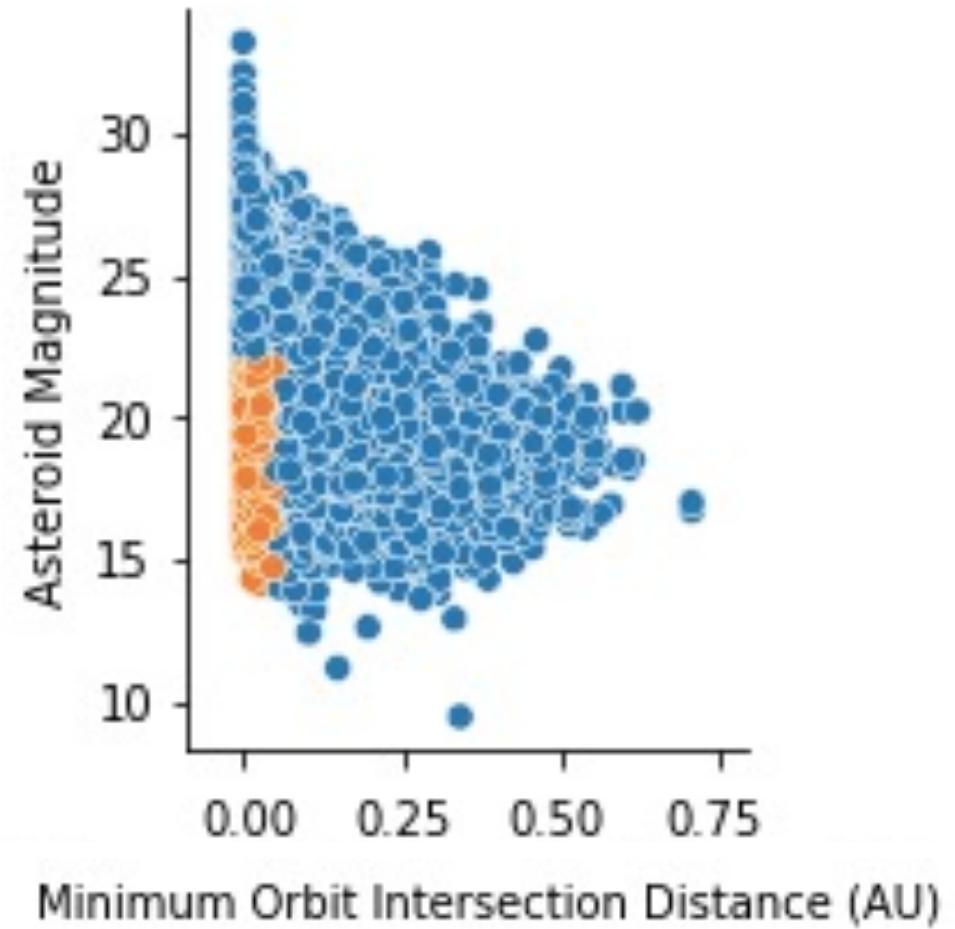
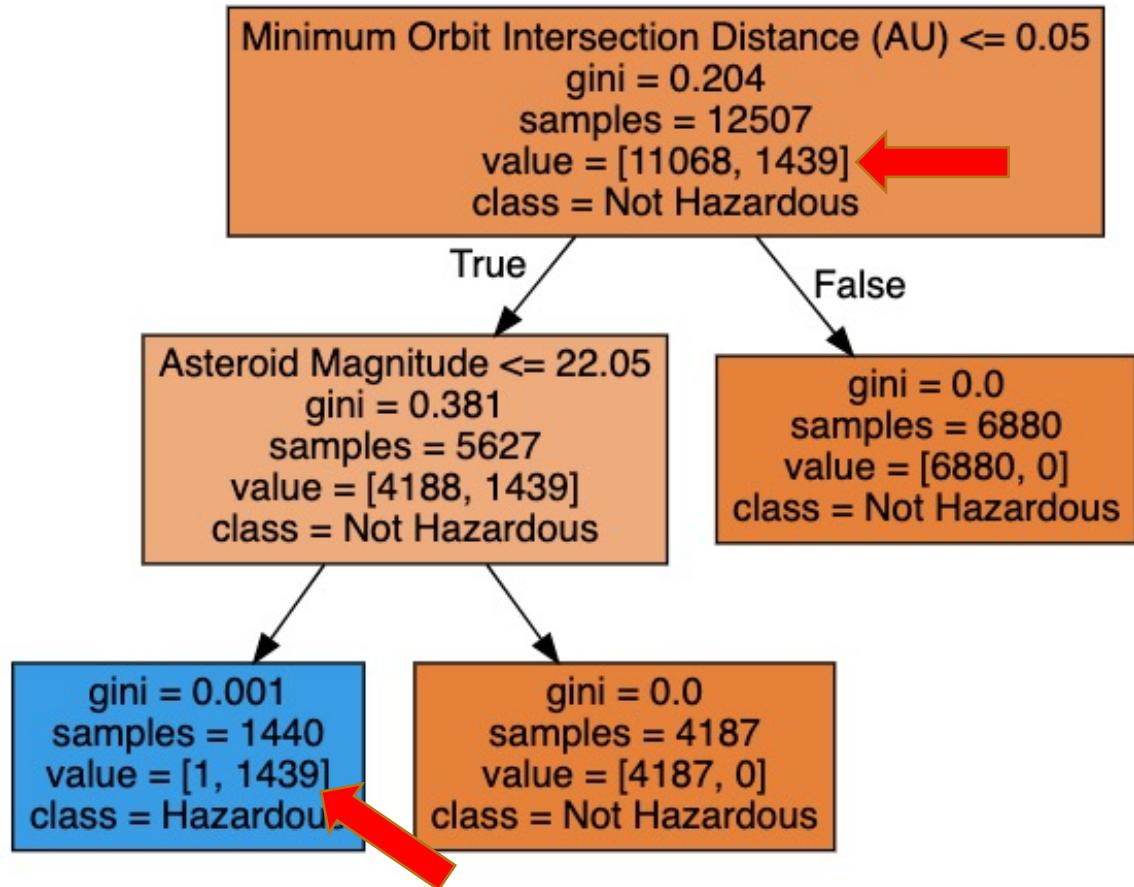


Neural Network Confusion Matrix



RESULTS

Model	Accuracy	Precision	Recall
K-Nearest Neighbors (with normalization)	0.891	0.333	0.006
K-Nearest Neighbors (without normalization)	0.891	0.5	0.006
Decision Tree	1	1	1
Naive Bayes	0.961	0.866	0.759
Logistic Regression(without normalization)	0.936	0.793	0.553
Logistic Regression(with normalization)	0.891	0	0
Neural Network	0.958	0.848	0.753



DECISION TREE

RESULTS

Model	Accuracy	Precision	Recall
K-Nearest Neighbors (with normalization)	0.891	0.333	0.006
K-Nearest Neighbors (without normalization)	0.891	0.5	0.006
Decision Tree	1	1	1
Naive Bayes	0.961	0.866	0.759
Logistic Regression(without normalization)	0.936	0.793	0.553
Logistic Regression(with normalization)	0.891	0	0
Neural Network	0.958	0.848	0.753

RESULTS

Model	Accuracy	Precision	Recall
K-Nearest Neighbors (with normalization)	0.889	0.467	0.168
K-Nearest Neighbors (without normalization)	0.892	0.6	0.018
Decision Tree	0.886	0.454	0.259
Naive Bayes	0.87	0.333	0.197
Logistic Regression (without normalization)	0.89	0.455	0.044
Logistic Regression (with normalization)	0.891	0	0
Neural Network	0.891	0	0

LESSONS

Data is everything

Analysis \geq ML

Be ready for unexpected results

Lessons

Accuracy is not always a good metric

Don't underestimate simple models

Be passionate and enjoy the journey

REFERENCES

- <https://www.kaggle.com/shrushtijoshi/asteroid-impacts>
- <https://www.kaggle.com/nasa/asteroid-impacts>
- <https://pandas.pydata.org/docs/index.html>
- <https://seaborn.pydata.org/tutorial.htm>
- <https://matplotlib.org/stable/api/index>
- <https://www.javatpoint.com/classification-algorithm-in-machine-learning>
- <https://www.cnn.com/videos/us/2018/07/25/rise-in-near-earth-asteroids-original.cnn>