Near-Earth-Objects: Are We Doomed?

An analysis by Jacqueline Tsodikova, Alejandra Magana, Robert Janke, Michael Albers, Fred Jambor

What are we analyzing?

Our group is researching the likelihood an asteroid or comet will potentially harm our planet. Using historical data, we will determine what thresholds are used to predict which NEOs are the most hazardous to Earth.

Why this topic?

A new movie on Netflix, *Don't Look Up*, was just released that had to do with a comet approaching Earth and scientists trying to warn the public about it. Although this movie is more about comedic humor, a comet or asteroid harming our planet is definitely something that could happen. We wanted to research and analyze the data to see how likely a NEO could harm us.

How many close approach objects will we have in the next decade?

On average, 200-400 space objects enter Earth's atmosphere every year, that's about one a day. But, these are usually small and not hazardous

Can we predict potentially hazardous objects in the future?

The probability of an object to be PHA is small, but the damages are definitely big. A comet that was 10 kilometers is what wiped out the dinosaurs.

Which NEOs are the most potentially hazardous?

Knowing exactly which NEOs are the most hazardous can help researchers track those more closely.

Description of Our Data Source

Jet Propulsion Laboratory
California Institute of Technology

Our data comes from the Center For Near Earth Object Studies (CENOS) a NASA center that computes highly accurate orbital data for thousands of asteroids and comets in our solar system.

CNEOS collects it's information from minor planet center which includes orbital data and physical properties like size and rotation rates.

Description of the Data Exploration Phase

Database:

We used ERD and Postgres to store and analyze our data.

Machine Learning:

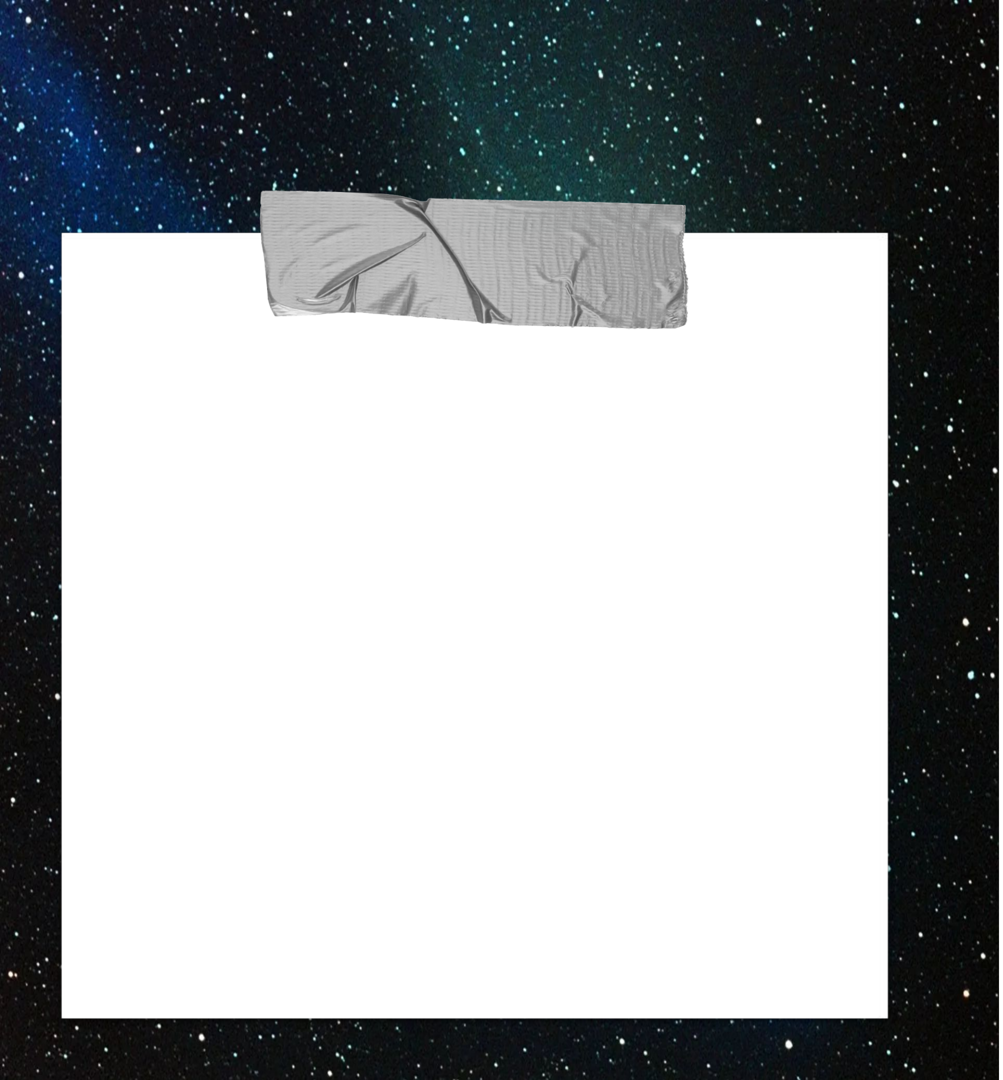
Analyzed NEOs that are potentially hazardous to Earth, by using the Random Forest Classifier, Over/Under Sampling, RandomOverSampler and SMOTE to test the accuracy of the data set.

Description of the Analysis Phase

The Random Forest Classifier had a 92% accuracy.

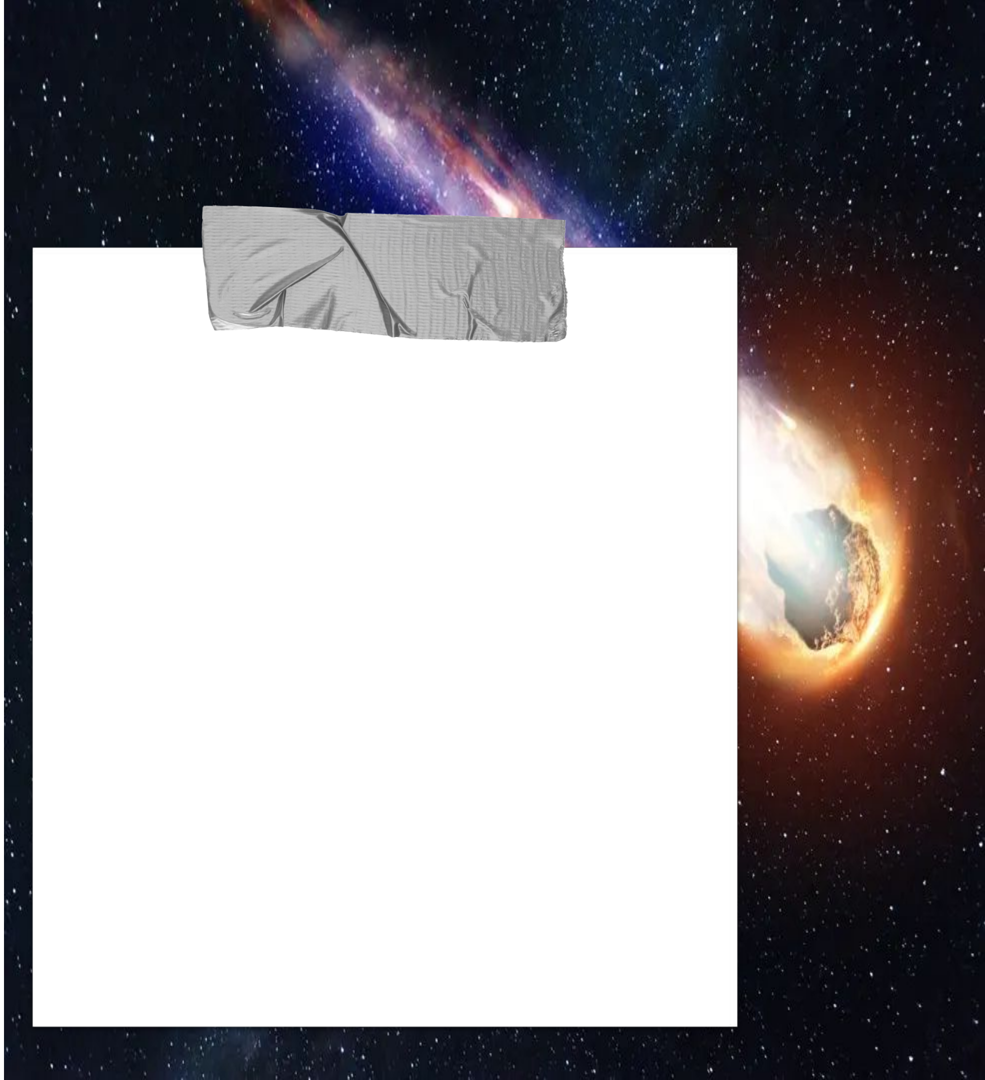
Over/Under Sampling had almost 40% of actual impacts classified wrong (228 predicted wrong, 325 predicted correctly.

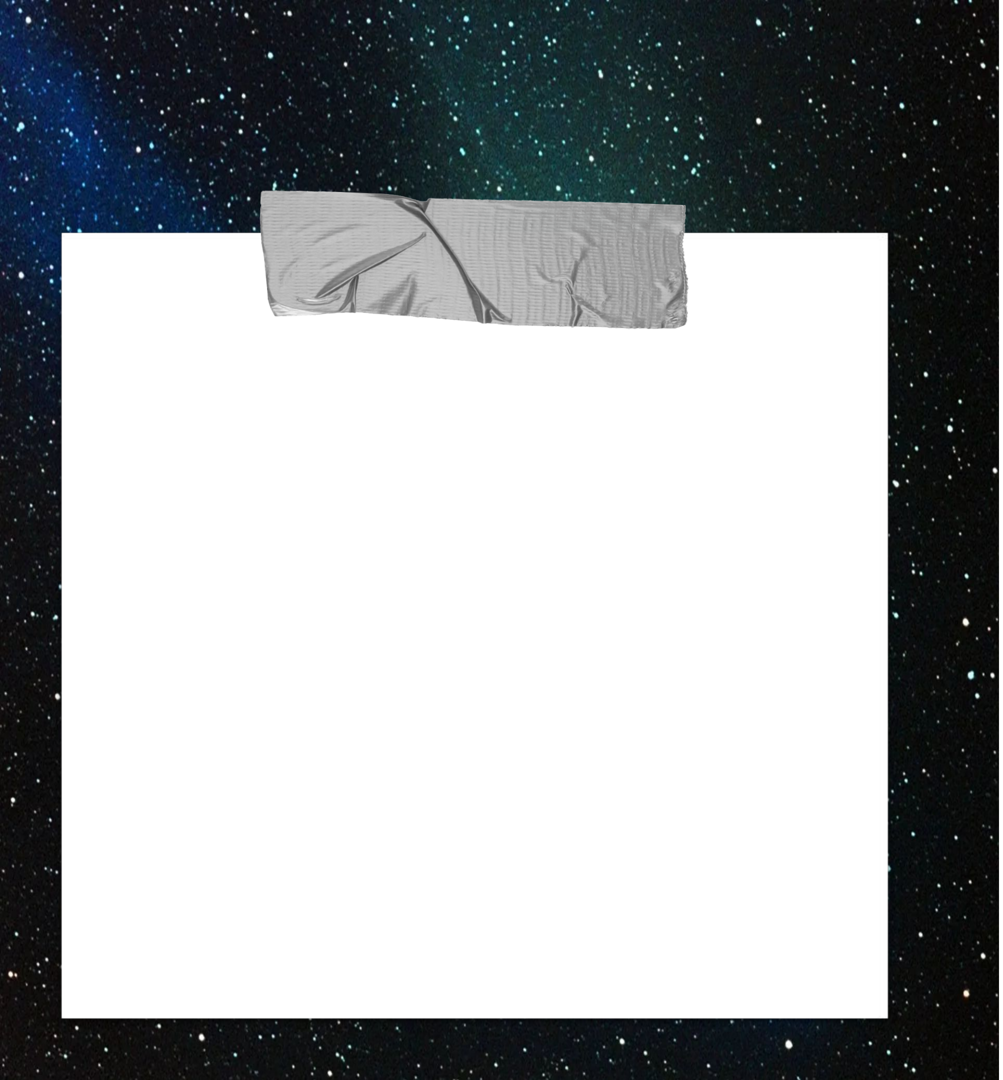
RandomOverSampler and SMOTE failed to have an accurate prediction of hazardous object but we knew it was good to test different models on our machine learning in case any data is skewed or gave more accurate results.



Storyboard

Description of the Interactive Elements





Description of the Tools for Our Dashboard

Description of Interactive Element(s)

Technologise Used:

- Tableau for our dashboard
- AWS to store data
- Pandas for cleaning the tables
- Random Forest Classifier to test our overall performance
- PostgreSQL to store large and sophisticated data safely
- PySpark to have a wide range of libraries
- Entity Relationship DIagram (ERD) to model the stored data

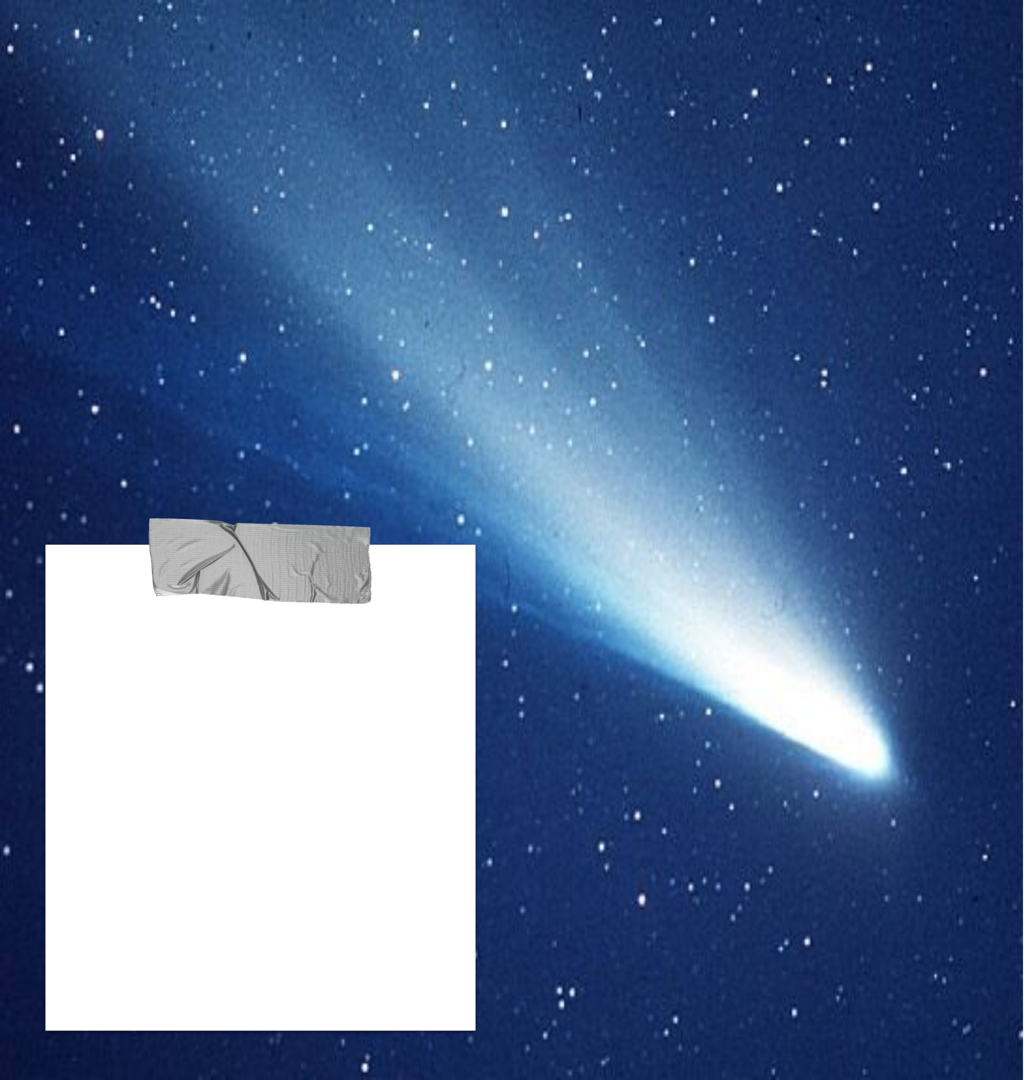
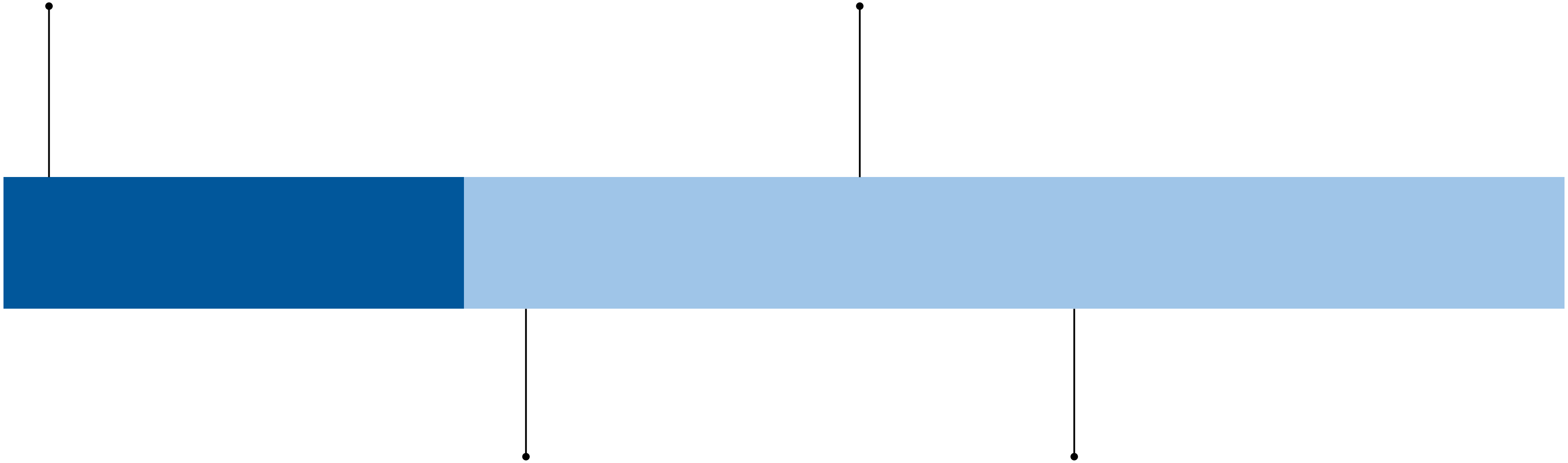


Tableau Story

-

Timeline



Analysis Results