Asteroid Size Predictor Web Application
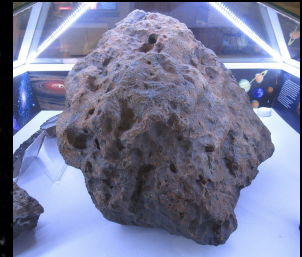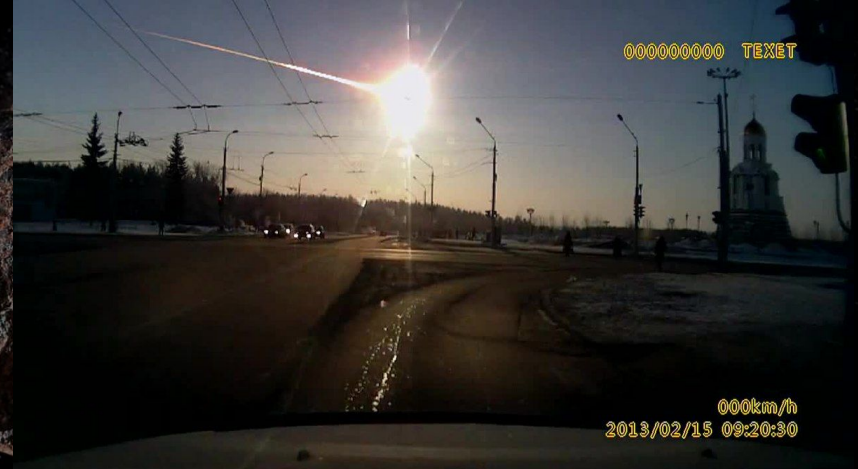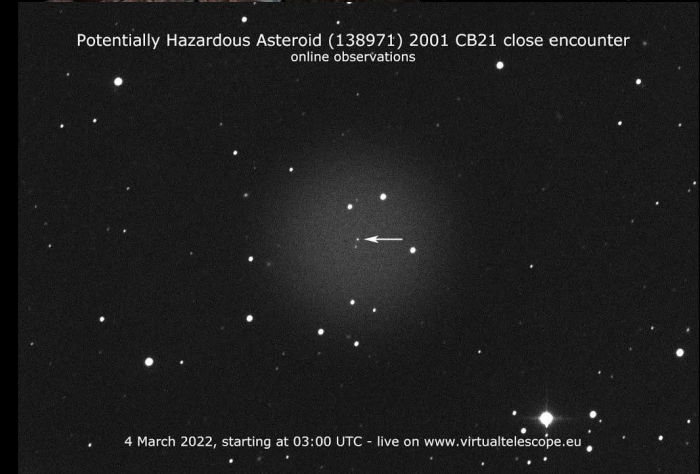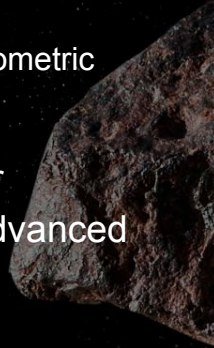
Nate DiRenzo

Introduction

Why Asteroids?

- There are over 1,000,000 known asteroids in our solar system

- A relatively small (140-300m) impact would have continent-wide effects

- Lots of public data available

Current State

- Asteroid sizes are typically estimated from thermal radiometry

  - Function of Absolute Magnitude, Geometric Albedo

- Precise measurements require stellar occultation and direct imaging with advanced instruments on Earth (not many)

Potentially Hazardous Asteroid (138971) 2001 CB21 close encounter
online observations

4 March 2022, starting at 03:00 UTC - live on www.virtualtelescope.eu

Project Goal

Build a machine learning model to predict asteroid diameter.

Determine efficacy of model.

Make model publicly available for others to use.
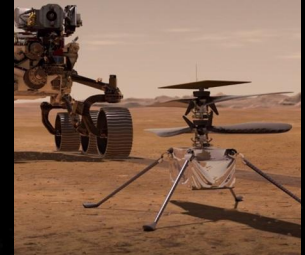
Solution Path

The Data

- [Small-Body Database](#)

  - Courtesy of NASA Jet Propulsion Laboratory, CalTech

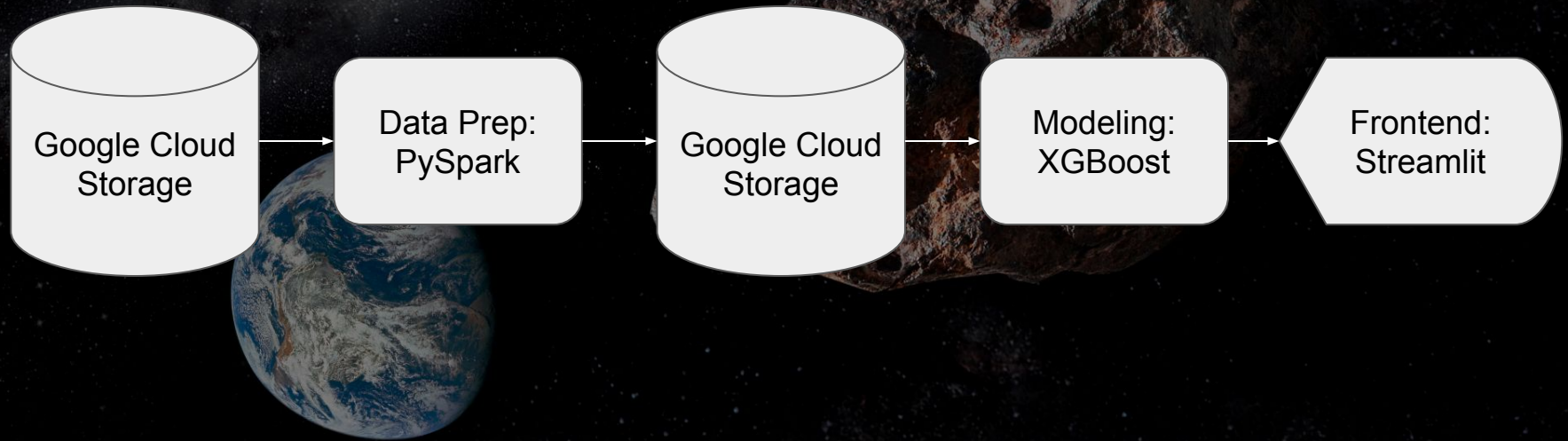- Contains 1.2 million + distinct asteroids, with 30+ features

JPL **Jet Propulsion Laboratory**
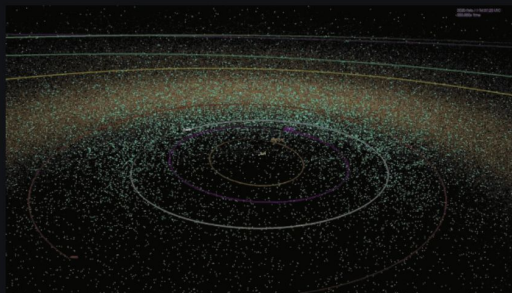California Institute of Technology

Caltech

Data Pipeline

Google Cloud Storage → Data Prep: PySpark → Google Cloud Storage → Modeling: XGBoost → Frontend: Streamlit

Streamlit Application

Conclusions

MAE: .979
R2: .79

Model Results

Areas for Future Work

- More nuanced data preparation & feature engineering

- Deploy PySpark model

- Improve Streamlit Application UI & Interactivity

- Provide functionality for predicting on new inputs

Thank you!

Appendix

Correlation Heatmap

Asteroid Diameter Distributions - Pre/Post Quantile Trim
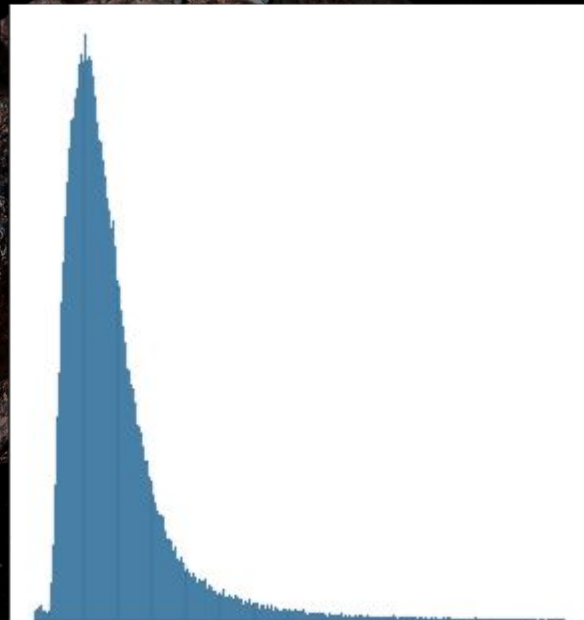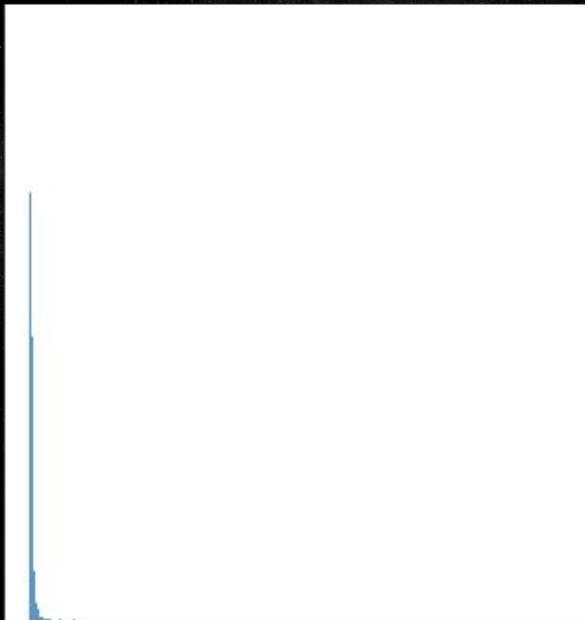
XGBoost Model Feature Importance