Potentially Hazardous Asteroid Classification

Laura Miller

Overview

- Asteroid impact probability low in the short-term, inevitable in the long-term
- Global extinction events: 26 million-year cycle
- Tsunguska events: better example of what we may have to contend with in the near future

Project output: Random Forest model that classifies whether or not an asteroid is potentially hazardous (binary classification problem), based on various features.

Outline

- Business Problem
- Methods
- Data
- Results
- Conclusions

Business Problem

NASA's NEO Observations Program is tasked with finding, tracking, and characterizing NEOs. Classification models can flag potentially hazardous asteroids as updates are made to the MPC (Minor Planet Center) so that the CNEOS (Center for Near-Earth Object Studies) can follow up by tracking their orbits and analyzing their possible future positions relative to Earth.

A classification model should prioritize recall/sensitivity for the positive class to ensure that any asteroid that is truly potentially hazardous is properly identified.

“I don’t want to be the embarrassment of the galaxy to have had the power to deflect an asteroid, and then not and end up going extinct. We’d be the laughingstock of the aliens and the cosmos if that were the case.”

Neil deGrasse Tyson

Methods

01 Scrub

Impute missing values, remove outliers on basis of feature weight/importance.

02 Explore

Extract insights, examine relationships between features and target.

03 Model

Try models (Logistic Regression, Decision Tree, Random Forest). Model selection and tuning.

04 Interpret

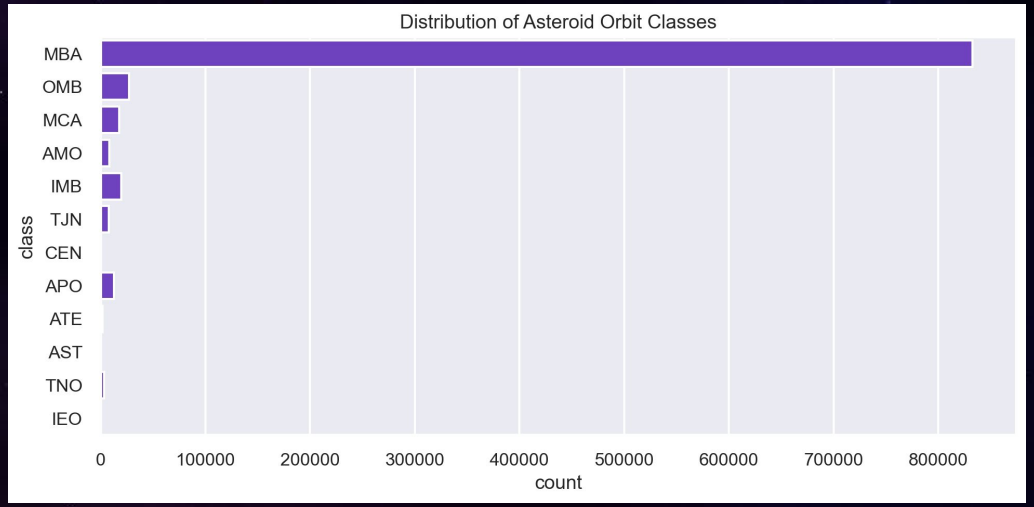
Evaluate best model performance and feature importances.

Data

Maintained by Jet Propulsion Laboratory of California Institute of Technology, under NASA

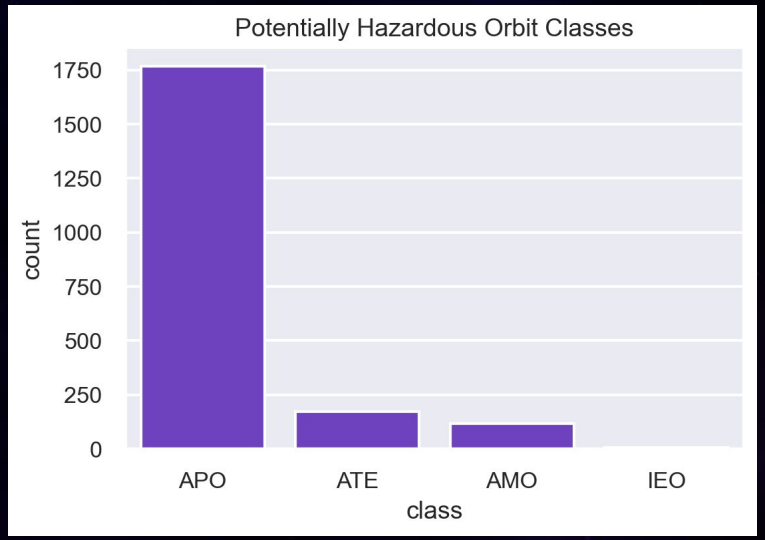
- JPL Small-Body Database Search Engine
  - 921,430 entries and 44 features (post-scrubbing, including dummies)
  - Target: ‘pha’ flag, whether or not asteroid is potentially hazardous
  - Class imbalance
    - 99.8% non-hazardous (negative)
    - 0.2% potentially hazardous (positive)

Data

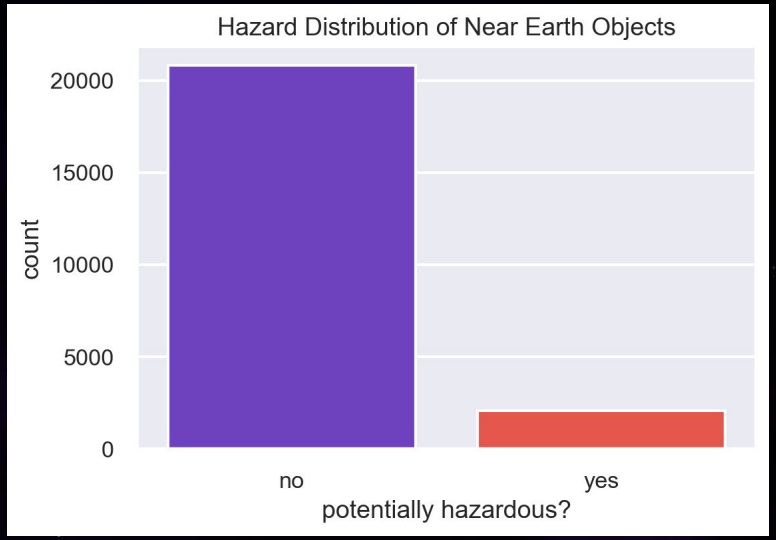


- Most asteroids are Main Belt Asteroids, which lie in the asteroid belt. Their orbital elements are constrained by (2.0 AU < a < 3.2 AU; q = 1.666 AU).
- IEO (Interior Earth Objects) class contains the fewest asteroids.

Data

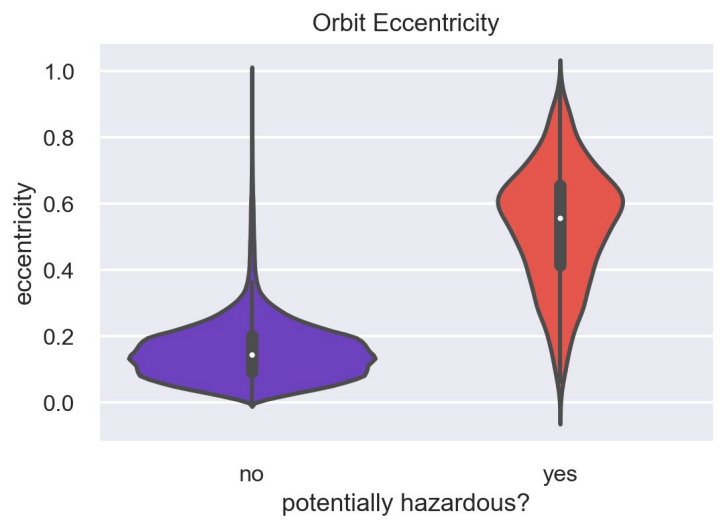


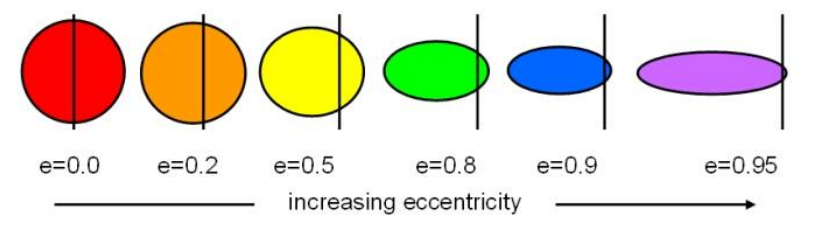
Most potentially hazardous asteroids are APO (Apollo class), with near-earth orbits similar to that of 1862 Apollo (a > 1.0 AU; q < 1.017 AU)



Of NEOs, 91% are non-hazardous, while 9% are potentially hazardous

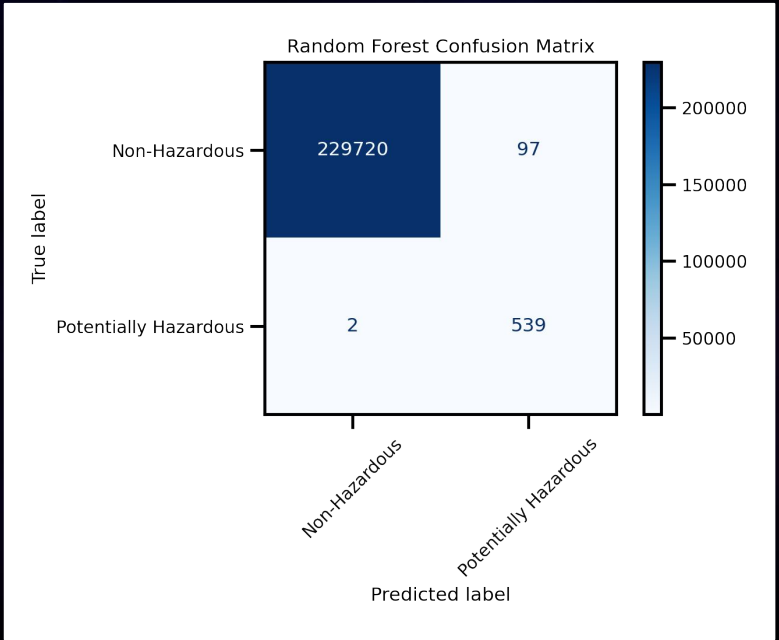
Data





- Potentially hazardous asteroids seem to have more elliptic orbits
- Orbits are more elliptic as e approaches 1
- Eccentricity could contribute to orbit velocity

Results



Best model: **Random Forest**

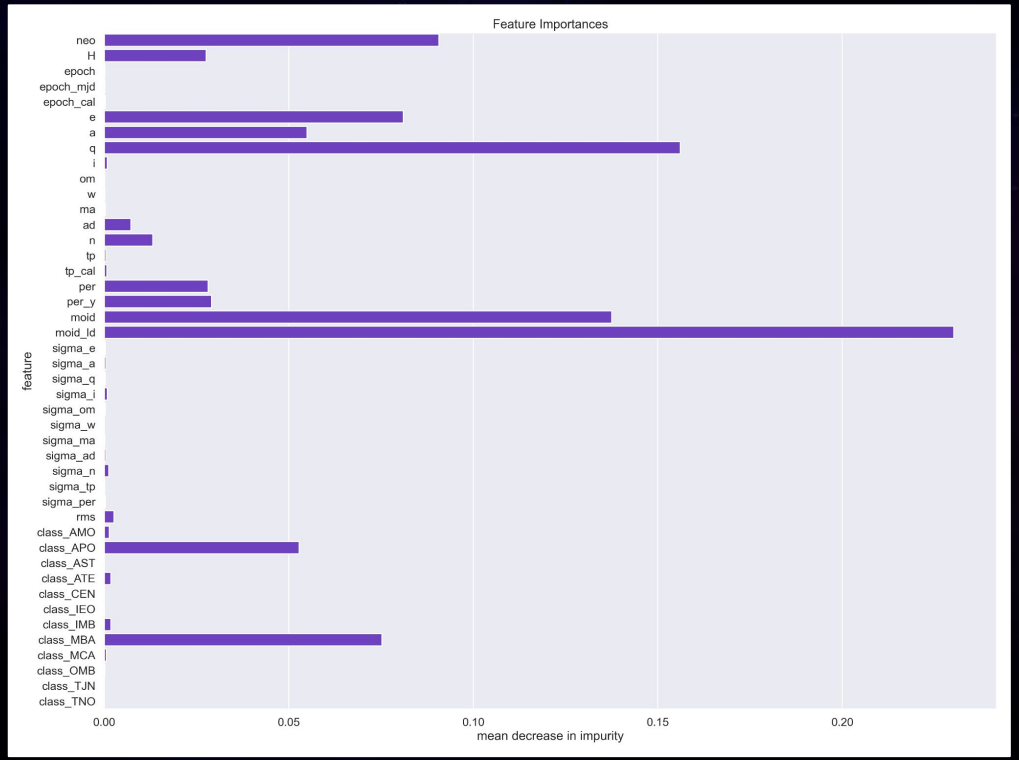
Macro Avg Scores:

- Precision: 0.92
- **Recall**: 1.00
- F1-score: 0.96

Confusion matrix (test performance):

- Missed 2 phas
- Misclassified 97 non-hazardous asteroids

Results



Important features:

- moid/moid_ld
- q (perihelion distance)
- neo
- e (orbit eccentricity)
- class_MBA

Conclusions

Built a model that prioritizes the classification of every potentially hazardous asteroid (high recall for the positive class).

- Next steps:
  - Search more feature weights
  - Retrain model as updates are made to JPL Small-Body Database
    - More asteroids with known labels
    - Better characterization of recorded asteroids

Thank You!

Email: laura@warrahdesign.com

Github: @NariMo91