Proposal: Under Attack - Binary Classification of Hazardous Objects from Space

Katie Knight*
Anna-Maria Nau*
Christoph Metzner*
kknigh22@vols.utk.edu
anau@vols.utk.edu
cmetzner@vols.utk.edu
DSE 511 - Project 3
University of Tennessee, Knoxville
Knoxville, Tennessee, USA

Figure 1: Potential threat of extinction through extraterrestrial objects. (https://eparisextra.com/living/nasa-attempts-to-stop-hypothetical-asteroid-from-hitting-earth-and-fails/)

ABSTRACT

This project applies binary machine learning classifiers to categorize potentially hazardous objects, like asteroids and comets, to serve as an early warning system. The origin of the dataset is from the Kaggle challenge “NASA: Asteroids Classification” (see https://www.kaggle.com/shrutimehta/nasa-asteroids-classification).

*All three authors contributed equally to this research.

Knoxville ’21, Fall, 2021, Knoxville, TN

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

KEYWORDS

dataset, machine learning, extraterrestrial object, classification

ACM Reference Format:
Katie Knight, Anna-Maria Nau, and Christoph Metzner. 2021. Proposal: Under Attack - Binary Classification of Hazardous Objects from Space. In *Knoxville ’21: DSE511 - Introduction to Data Science and Computing I, Fall, 2021, Knoxville, TN*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

In "A Synopsis of the Astronomy of Comets", Edmond Halley first voiced his concerns about the devastating impact of extraterrestrial objects, such as asteroids or comets, on the modern world in the 18^{th} century [8]. A group centered around Luis Alvarez hypothesized that such objects caused several mass extinctions and, generally, has a significant influence on the well-being of our planet [1]. In the face of potential extinction, US Congress ordered NASA, in collaboration with The University of Arizona and Hawaii, to develop strategies for observation, tracking, characterization, and damage mitigation [5] for such objects. One specific strategy is the development of statistical models utilizing machine learning that are able to classify potentially hazardous objects based on their properties, such as shape, size, velocity, or shape of orbit.

These extraterrestrial objects, asteroids or comets, are referred as near-earth objects (NEOs). NEOs orbit the sun, may potentially cross Earth's orbit [2], and can range in size from a dust particle to kilometers in diameter [4]. There is a wide range of publications that propose different strategies for impact risk assessment of NEOs based on common features such as size or velocity. For example, a roughly twenty-year-old approach is to use scales rating the potential impact of NEOs such as the *Torino scale* or *Palermo scale* [6]. Their advantage is that the layman can easily interpret such scales. Another approach focuses on using VNIR spectroscopy to characterize the compositional and physical structure of NEOs to determine their hazardous potential [3].

[7] were the first to apply machine learning algorithms to the classification of hazardous NEOs. A more recent study successfully applied more complex machine learning algorithms based on neural network frameworks to infer the physical properties and impact risks of asteroids by analyzing energy deposition curves [9]. We hypothesize that using machine learning algorithms can play a vital part in the detection of hazardous objects. Specifically, we ask if standard "off-the-shelf" machine learning algorithms can achieve outstanding performance classifying such objects.

2 DATA

This binary classification analysis was conducted using the NASA: Asteroids Classification dataset, collected from the NASA API, available at https://www.kaggle.com/shrutimehta/nasa-asteroids-classification. This dataset contains the information of 4687 asteroids, where each asteroid represents one sample. The majority of asteroids, 3932 samples, are of non-hazardous in nature, whereas 755 were labeled as hazardous. This label imbalance has to be thought of during the classification process. The dataset has 39 features and one column containing the ground-truth labels. All features but one contain numerical values with different magnitudes. The dataset includes typically measured features of an asteroid, such as the *absolute magnitude*, *estimated diameter*, *relative velocity*, distance measures of the object to the sun such as *perihelion distance* (minimum distance) and *aphelion distance* (maximum distance), or the shape of the objects orbit represented by *eccentricity*. The final report will list a detailed description of all features utilized in the supervised classification process. Due to redundancy or lack of diversity in information some features will be removed during the pre-processing task. For example, features describing the *estimated diameter* of the asteroids in different units (e.g., feet, meter, and miles) will be dropped due to redundancy. The original dataset will be randomly split into a training set containing 80% and a testing set representing 20% of the original data.

3 METHODS

This work focuses on binary-supervised classification of extraterrestrial objects using machine learning algorithms. We will use and compare four different algorithms: 1) naive Bayes as the baseline, 2) support vector machine, 3) decision tree, and 4) k-Nearest Neighbors. Furthermore, we want to investigate how dimensionality reduction via principal component analysis impacts classification performance. The hyperparameters of all algorithms will be optimized using k-fold cross validation (k=5) on the training set via the function *GridSearchCV* from the machine learning library *sklearn*. The performance of the developed models will be evaluated on the testing set using the widely accepted performance metrics *F1-Score*, *Precision*, and *Recall*. Success will be measured by whether hazardous asteroids are correctly identified.

4 CONCLUSION

We want to demonstrate that machine learning algorithms are a vital method in identifying extraterrestrial objects as hazardous using descriptive measurements. Furthermore, we want to see which type of algorithm performs best on this type of data and if dimensionality reduction has a positive or negative effect on the performance of the models. The impact of this work is to demonstrate how simple, "off-the-shelf" machine learning models can be a powerful tool in the vital task of protecting our only planet earth from potentially life-ending threats from space.

REFERENCES

[1] Luis W Alvarez, Walter Alvarez, Frank Asaro, and Helen V Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* 208, 4448 (1980), 1095–1108.
[2] Space Studies Board, National Research Council, et al. 2010. *Defending planet earth: Near-Earth-Object surveys and hazard mitigation strategies*. National Academies Press.
[3] Michael J Gaffey, Paul A Abell, and Paul S Hardersen. 2006. COMPOSITIONAL AND PHYSICAL CHARACTERIZATIONS OF NEOs FROM VNIR SPECTROSCOPY. (2006).
[4] Amy Mainzer, T Grav, J Bauer, J Masiero, RS McMillan, RM Cutri, R Walker, E Wright, P Eisenhardt, DJ Tholen, et al. 2011. NEOWISE observations of near-Earth objects: preliminary results. *The Astrophysical Journal* 743, 2 (2011), 156.
[5] NASA. [n. d.]. Near-Earth Object Observations Program. https://www.nasa.gov/planetarydefense/neoo.
[6] NASA. [n. d.]. Torino Impact Scale. https://web.archive.org/web/20070224184143/http://impact.arc.nasa.gov/torino.cfm.
[7] Carrie R Nugent, John Dailey, Roc M Cutri, Frank J Masci, and Amy K Mainzer. 2017. Machine learning and next-generation asteroid surveys. In *AAS/Division for Planetary Sciences Meeting Abstracts# 49*, Vol. 49.
[8] Alfred Romer. 1984. Halley's comet. *The Physics Teacher* 22, 8 (1984), 488–493.
[9] A. M. Tarano, J. Gee, L. Wheeler, S. Close, and D. Mathias. 2020. Automating the Inference of Asteroid Physical Properties and Motion. In *AGU Fall Meeting Abstracts*, Vol. 2020. Article P008-02, P008-02 pages.