High-Risk Asteroid Impact Classification by Machine Learning Algorithms

Submitted By: Pranav Meraga, Chanwook Park

Dept. of Mechanical Engineering

Northwestern University

Date Submitted: 08/28/22

MDS Summer Course 2022

Table of Contents

ABSTRACT………………………………………………………………………………… **pg. 3**

INTRODUCTION……………………………………………………………………… **pgs. 3-6**

Burst Altitude Derivation………………………………………………….……. **pgs. 4-5**

K-Means Clustering……………………………………………………………... **pgs. 5-6**

RESULTS…………………………………………………………………………….... **pgs. 6-7**

CONCLUSION………………………………………………………………….… **pgs. 7-8**

CODE…………………………………………………………………………………….. **pg. 8**

ACKNOWLEDGEMENTS……………………………………………………………….. **pg. 8**

REFERENCES………………………………………………………………………….. **pg. 8**

High-Risk Asteroid Impact Classification by Machine Learning Algorithms

Pranav Meraga

ABSTRACT: Since its creation in 1998, NASA's Center for Near-Earth Object Studies(CNEOS) has detected well over 28,000 potentially hazardous asteroids and comets that orbit within a close proximity to Earth. However, almost all of these objects orbit are either at distances far beyond Earth's gravitational reach or are to small to enact severe damage to be considered an imminent threat. Despite these classifications, incidences have still occured with one notable miss call being in 2013 where a 20-meter asteroid exploded over the surface of Chelyabinsk, Russia. Over 1600 people were injured due to the airburst and the only possible sightings of the event were captured by the Suomi NPP satellite 3 hours later. Even with a tenured program in place to prevent such anonymities, the tremendous room for error has allowed such catastrophes occur, but never again. Since the accident, NASA's prediction software, Sentry, has mapped out many orbits of thousands of NEOs, and the research below is dedicated to further improving of NEO classification using optimized categories for the purpose of high-risk analyses to prevent potential mishaps like before. Currently, Sentry classifies hazard levels of NEOs with 40+ attributes including but not limited to Eccentricity, Inclination, Orbital Period, etc.; however, this is a very tedious classification process that also doesn't account for the worst case scenario(like Chelyabinsk). In the following research paper, by using a simple K-Means clustering algorithm, a full dataset of over 4,000 asteroids can be categorized into 3 distinct hazard groups based entirely on 3 key attributes: Velocity Relative to the Earth(km/s), Estimated Diameter(km), and Burst Altitude(km). The program decided that 3 clusters was necessary to characterize the data based on the 3 dimensions listed above. Even though the algorithm perfectly scoped out the most hazardouss of the asteroid dataset, it did not associate at all with NASA's set classifications at all and therefore is not the most trusted approximation due to many lacking factors(Asteroid Porosity, Orbital Distance, etc.). In the future, it would be most optimal to include additional information to draw accurate conclusions.

KEYWORDS: *Airburst, Sentry, Eccentricity, K-Means Clustering, Porosity*

INTRODUCTION

To clarify, asteroids will be the primary focus of the research at hand. While both asteroids and meteoroids are rocky, metallic objects in space, meteoroids range between a small particle to 10-meters in length and have minimal effect on the Earth's atmosphere in a collision(which is a daily occurence). While small particles may collide with Earth everyday, the possibility of a devastating impact is still likely, with "An object similar in size to the Chelyabinsk meteor is set to impact Earth every 10-100 years."(*Phys.org)*. The frequency of impacts is an astonishing level of consistency that can pose severe threats if not considered thoroughly. The essential purpose of this research project is to classify these Potentially-Hazardous Objects(PHOs) under the most scrutinous circumstances within 3 dimensions(Velocity, Diameter, and Burst Altitude). As stated before, NASA's CNEOS has a public database for asteroid detection that was used for the basis of this research. The dataset, found on Kaggle and complied by NASA, contains +4,000 asteroids

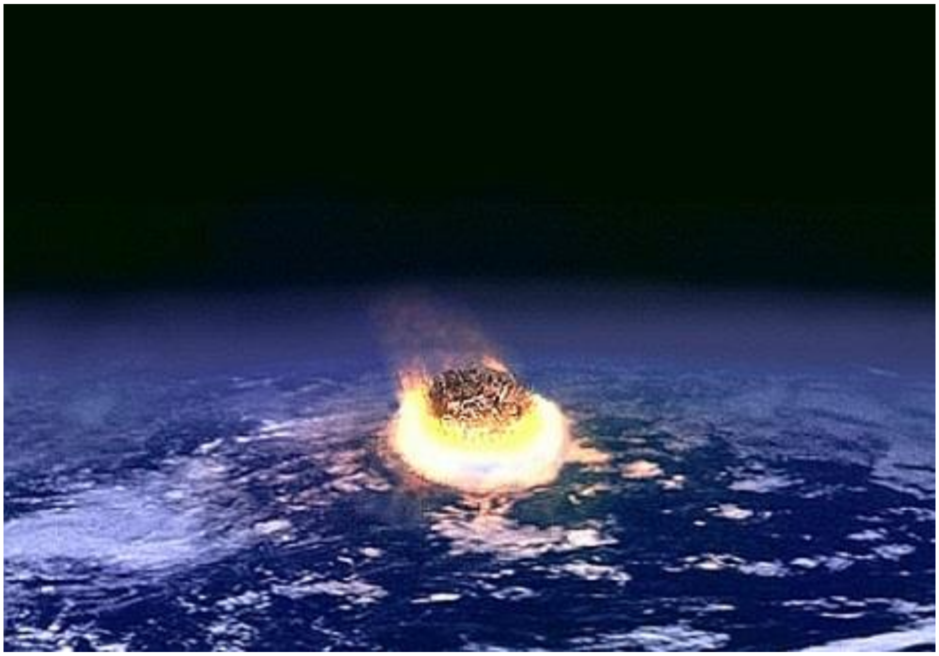

Figure 1. An Impact Event, Source: http://en.wikipedia.org/wiki/Impact_event

that are classified as either hazardous or not based on over 40 parameters. Using only 2 out of the 40 columns(as well as another value derived in the following section), a new synopsis of asteroid potential hazard level can be made under the most devastating circumstances.

Burst Altitude Derivation

Airbursts are meteors that combust due to an influx of compressed air whilst entering the atmosphere and only leave a minute percentage of debris as meteroites on the surface. Of the 40 features NASA considers in hazard detection, the altitude in which an asteroid bursts is not included, even for the asteroids they do deem as hazardous. The level in which an asteroid explodes is vital in determining its impact on the people and the surrounding environment. For example, the Chalyabinsk Meteor injured 1600 Russian civilians solely due to the shockwaves of the explosion that broke windows and car windshields. In properly risk assessing based on high mortality, the altitude in which an asteroid combusts is vital in determining the amount of debris that can reak havoc in the off-chance of collision.

$$z_{b,1\%} = 13 - 6.04\log_{10}E_{Mt} - 0.88\left(\log_{10}E_{Mt}\right)^2$$

$$z_{b,50\%} = 25.7 - 7.83\log_{10}E_{Mt} - 0.31\left(\log_{10}E_{Mt}\right)^2$$

$$z_{b,99\%} = 47.9 - 8.43\log_{10}E_{Mt} - 0.03\left(\log_{10}E_{Mt}\right)^2$$

Figure 2. Logorathmic Equations that equate Burst Altitude(km) using kinetic energy(Mt). Zb,50% is the average calculation value of the range of possible combustion altitudes

The influx of compressed gas within an asteroid is caused by the porosity, the amount of holes in the material, of the asteroids composition. That along with the asteroid’s permeability, the process of burning the rocky outer layer due to air resistance, causes the asteroid to disintegrate into an airburst. In a 2017 study by Purdue University, the scientists were able to derive that by deriving the entrance energy of the system(K = ½(M*V^2)), one can predict the approximate longitude of when an asteroid will combust. From the given dataset, the entrance velocity can be determined solely on 2 given parameters: Estimated Diameter and Relative Velocity. One key issue in this derivation is how to determine the mass component. This can be solved by generalizing that all the asteroids are made of stony material(with minute percentages of iron) which is about 95% of all classified asteroids. Using this generalization, one can associate a common density value of around 3400 kg/m^3, as well as conform an asteroid’s mean shape to be spherical and multiply density with the volume of a sphere to get the overall mass. After using these three values for each asteroid, one can determine the burst altitude within these parameters.

K-Means Clustering

After configuring the Burst Altitude as a metric, it was a matter of how to categorize the dataset into levels of hazard based on the 3 variables. Therefore, data clustering is required to fit a 3-dimensional dataset such as this to differentiate specific values of each attribute and how they fit under a specific hazard classification. Since it is established that the dataset is measured using 3-dimensions, it is a matter of settling with a fixed number of clusters. K-means clustering is essential in finding the efficient number of groups that can effectively differentiate the data without overfitting(inefficient and excessive groupings of data).

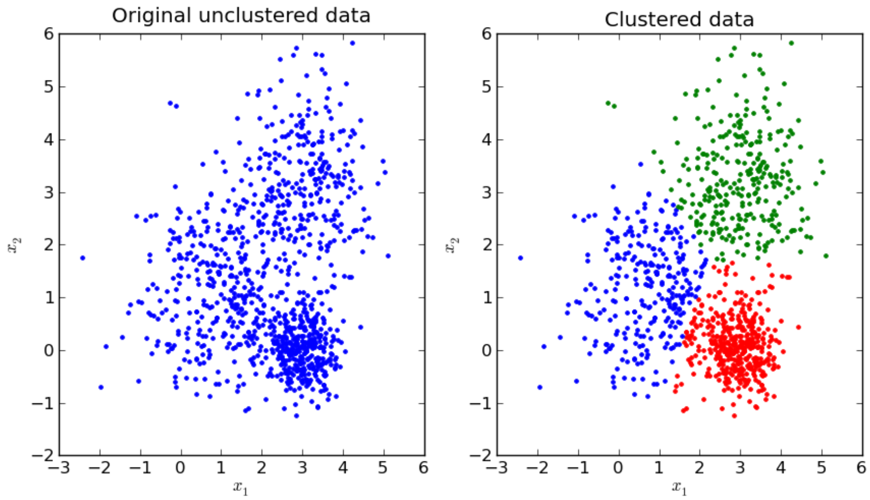

Figure 3. K-Means Clustering based on Machine Learning to group sets of datapoints

Machine learning helps find the best quantity of clusters using a repetitive cycle of learning and reconfiguration over a K-amount of clusters. After each iteration, the GVF or general variance of clusters is measured using the Sum of Squares from a center value that is randomly determined after repetitive trial and error per the number of clusters provided. The GVF will increase over the number of clusters, till it is 1 datapoint per cluster for full variance, which is very inefficient. By graphing the GVF at each number of clusters, the change in slope of GVF(the "elbow" of the graph) is where the most optimal cluster value lies before processing into overfitting. This methodology is very rudimentary, but allows machine learning to normalize the data for efficient diagnoses and modeling within a dataset.

RESULTS

In the case of the asteroid dataset, the GVF graph produced a 'bend' at around 3 clusters to produce optimal categorization of the values based on velocity, diameter, and burst altitude.

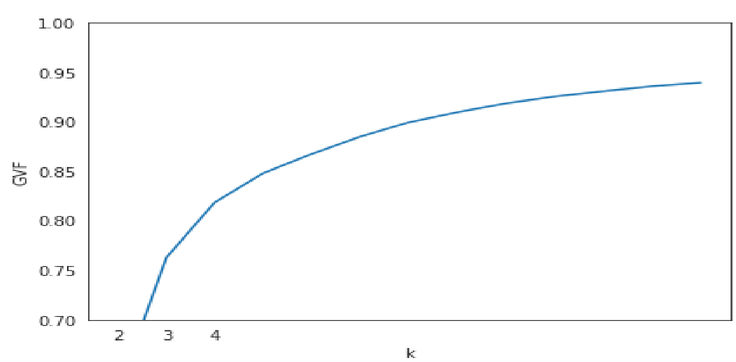

Figure 4. A Generalized Variance Function Graph of the Asteroid Data in relation to the k-values of clusters.

The above characterization produced a 3D rendering of the dataset into the 3 marginalizations of asteroid hazard levels using their respective features.

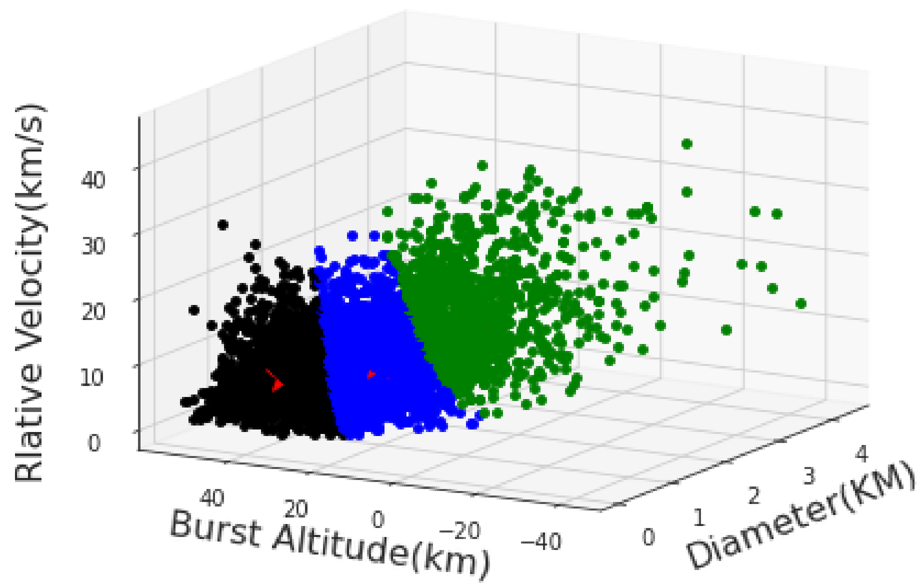

Figure 2. A 3 dimensional graph with Relative Velocity(km/s) in the z-axis, Burst Altitude(km) in the y-axis, and Diameter(km) in the x-axis.

As shown, the black cluster lies around the 0-1 km range for diameter, at high burst altitudes, and has entrance velocities in the lower end of the spectrum. These attributes of asteroids leads to a non-hazardous classification within that cluster. For the green cluster, the diameter lengths increased to atmost 4 km in diameter with a negative airburst distance, an indication of extreme hazard levels if in contact with Earth. The intermediary cluster wavers between hazardous and not with its range of 1-2.5 wide asteroids at relatively low burst altitudes and moderate velocities opens a plethora of classes of hazard at the respective ends of the range.

CONCLUSION

In conclusion, the data processes of this research outputted inferable results that satisfied the purpose of the project, but lacks in feasible applicability. The research at hand focuses on analyzing asteroid impacts in scrutinous situations of high risk propagation. Therefore, the possibility of application holds no basis of realism due to the high imporbability of impact projected in the CNEOS established asteroid watch. Not to mention, almost ⅔ of the dataset proves to have severe outcomes in the off chance of impact, a conclusion that seems grim but is incoherent with realistic approximations based on imperative variables such as orbit estimation and epoch date. Nonetheless, this research was still a necessary precaution in hindsight due to the major mishap the CNEOS miscalculated in 2013 Chelyabinsk, Russia that could potentially occur once more in the near future. Overall, high-risk analysis is a necessary, yet limited study for major catastrophes like asteroids, but should be more considerate of impactful variables in the future to get the most approximate results in even the most treacherous outcomes.

CODE: K-Means Clustering Algorithm for Asteroid Data

ACKNOWLEDGEMENTS

Special thanks to Wing Kam Liu and the rest of the PHD students at the Northwestern Mechanistic Data Science Summer Course for entrusting a thorough education of data science applications in mechanical engineering. Thanks to Chanwook Park for being an invaluable advisor and resource in helping generate the machine learning algorithm for this project.

REFERENCES

- Codebasics. 2022. *Best platform to learn technology in a most intuitive way*. [online] Available at: <https://codebasics.io/> [Accessed 3 September 2022].
- *A numerical assessment of simple AIRBLAST models of impact airbursts*. (n.d.). from https://onlinelibrary.wiley.com/doi/10.1111/maps.12873
- *Asteroid impact*. Real World Physics Problems. (n.d.). Retrieved https://www.real-world-physics-problems.com/asteroid-impact.html
- Smith, Kimberly Ennico. "Tunguska Revisited: 111-Year-Old Mystery Impact Inspires New, More Optimistic Asteroid Predictions." *Phys.org*, Phys.org, 27 June 2019, https://phys.org/news/2019-06-tunguska-revisited-year-old-mystery-impact.html.
- *Air Penetration Enhances Fragmentation of Entering Meteoroids*. https://onlinelibrary.wiley.com/doi/10.1111/maps.13034.
- "How Large Asteroids Must Be to Destroy a City, State, Country, or the Planet." *Business Insider*, 30 June 2018, https://www.businessinsider.in/slideshows/miscellaneous/how-large-asteroids-must-be-to-destroy-a-city-state-country-or-the-planet/slidelist/64806811.cms#slideid=64806820.
- Register, Paul J., et al. "Asteroid Fragmentation Approaches for Modeling Atmospheric Energy Deposition." *Icarus*, Academic Press, 18 Nov. 2016, https://www.sciencedirect.com/science/article/pii/S0019103516301853#fig0003.
- "Russian Meteor Characteristics." *Math Encounters Blog*, 8 Sept. 2016, https://www.mathscinotes.com/2013/02/russian-meteor-characteristics/.
- "Palermo Technical Impact Hazard Scale." *Wikipedia*, Wikimedia Foundation, 11 June 2022, https://en.wikipedia.org/wiki/Palermo_Technical_Impact_Hazard_Scale.