Machine Learning Techniques for Predicting Potentially Hazardous Near-Earth Objects

Nataly Jimenez Cruz  
973-914-4282  
njimene1@stevens.eduDepartment of Computer Science  
Stevens Institute of Technology  
  
  
Dr. Kevin Lu  
klu2@stevens.eduDepartment of Electrical and Computer Engineering  
Stevens Institute of Technology

*Abstract*— This research utilizes a curated ten-year dataset of near-Earth object (NEO) close approaches, applying machine learning to predict potentially hazardous asteroids (PHAs). With 34,725 NEOs monitored, the study aims to improve impact risk assessments and early warning systems, enhancing threat mitigation from small solar system bodies through comprehensive data and advanced models.

# Introduction and Literature Search

The threat of celestial impacts, more specifically Near-Earth Objects (NEOs) began with the hypothesized ancient planet, Theia. After colliding with Earth 4.5 billion years ago, defending our planet has been a responsibility of NASA and many other organizations around the world. To this day, researchers are continuing to find ways to efficently and effeceively predict hazardous objects in order to protect our planet. With the use of machine learning, this is possible.

There a bunch of variables and features invovled that can help make these accurate predictions, some which include orbital data, size, and velocity. The Center for Near Earth Objects (CNEOs) and NASA contain unique data of NEOs as well as APIs in order to extract this information.

# Problem Description

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file “MSW\_A4\_format”.

Identify applicable funding agency here. If none, delete this text box.

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

# Research Conducted

## Data Collection

The dataset was curated by first extracting all Earth close-approach data for NEOs from the Small-Body Database (SBDB) Close-Approach Data API for dates between January 1, 2004 and June 1, 2024. Using the request library, a HTTP GET request was sent to the specific API URL retrieving data from the API. Because the response from the API was in JSON format, the loads() function was used to parse the string and convert it into a Python dictionary. A data frame was created using 11 fields as columns and a list of NEO data as rows.

* **des** - primary designation of the asteroid or comet
* **orbit\_id** - orbit ID used for the close-approach computation
* **jd** - time of close-approach (JD Ephemeris Time, TDB)
* **cd** - time of close-approach (formatted calendar date/time, TDB)
* **dist** - nominal approach distance (au)
* **dist\_min** - minimum (3-sigma) approach distance (au)
* **dist\_max** - maximum (3-sigma) approach distance (au)
* **v\_rel** - velocity relative to the approach body at close approach (km/s)
* **v\_inf** - velocity relative to a massless body (km/s)
* **t\_sigma\_f** - 3-sigma uncertainty in the time of close-approach (formatted in days, hours, and minutes; days are not included if zero)
* **h** - absolute magnitude H (mag)
* **fullname** - formatted full-name/designation of the asteroid or comet (optional - only output if requested with the fullname query parameter)

The last step was to write a python script that fetched specific orbital data of each object from one of NASA’s Open APIs, Asteroids – Near Earth Object Web Service (NeoWs), enhancing the original dataset with additional features for accurate hazardous prediction and visualizing orbits of celestial bodies. The orbital elements of each object could be viewed through Postman using the API URL and a NASA API Key. It is recommended to use the portion of an object’s ‘fullname’ found within parenthesis to extract data from NeoWs.

One of the biggest challenges with using this API or any NASA API was the limit of API requests available (1,000 request per hour). Because each NEO had to be looked up, a request was made for each object, totaling around 15 hours for complete data extraction. Time and progress libraries were used for handling sleep intervals and displaying progress bars.

The following orbital features were extracted from the NeoWs:

* **estimated\_diameter** – estimated diameter of the NEO. A dictionary containing minimum and maximum diameters in kilometers
* **data\_arc\_in\_day –** number of days spanned by the data-arc
* **observations\_used –** number of recorded observations of this orbit
* **orbit\_uncertainity –** MPC “U” parameter: orbit uncertainty estimates 0-9 with 0 being good and 9 being highly uncertain
* **minimum\_orbit\_intersection** – Earth Minimum Orbit Intersection Distance (MOID) in AU
* **jupiter\_tisserand\_invariant** – Jupiter Tisserand Invariant
* **epoch\_osculation** – when these orbital elements were determined, in seconds from the epoch
* **eccentricity** – eccentricity of the orbit
* **semi\_major\_axis** – semi major axis of the orbit in au
* **inclination** – inclination of the NEO’s orbit in degrees
* **ascending\_node\_longitude** – longitude of the ascending node in degrees
* **orbital\_period** – orbital period in days
* **perihelion\_distance** – perihelion distance in au
* **perihelion\_argument** – argument of perihelion in degrees
* **aphelion\_distance** – aphelion distance in au
* **perihelion\_time** – time of perihelion passage in Barycentric Dynamical Time (TDB)
* **mean\_anomaly** – mean anomaly in degrees
* **mean\_motion** – mean motion in degrees per day
* **is\_potentially\_hazardous\_asteroid** – returns True or False

## Data Preprocessing and Feature Engineering

The result from the data extraction phase was a dataset containing 15,355 rows of NEOs, 31 columns of features and 1 label, ‘is\_potentially\_hazardous\_asteroid’. Rows with one or more not a number (NaN) values were dropped from the dataset. The various types and statistical descriptions of the data frame were observed, and columns were renamed for readability. Additionally, a new column was created, combining the average between ‘max\_diameter\_km’ and ‘min\_diameter\_km’ to create ‘avg\_diameter\_km’. The previous columns were dropped. Next, columns that were irrelevant for prediction were also dropped from the data frame. A label encoder function was used to convert the label column to numerical format i.e. 0 or 1.

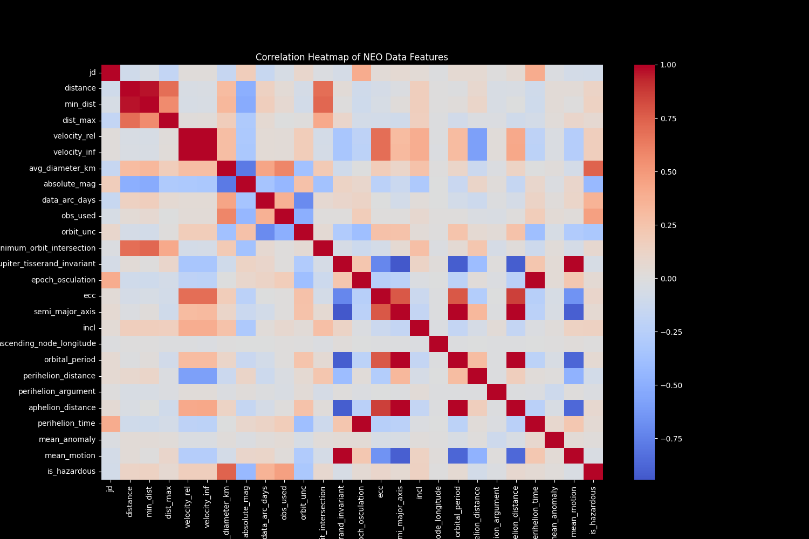


Fig. 1. Caption

A screenshot of a computer

Description automatically generated

Fig. 1. Caption

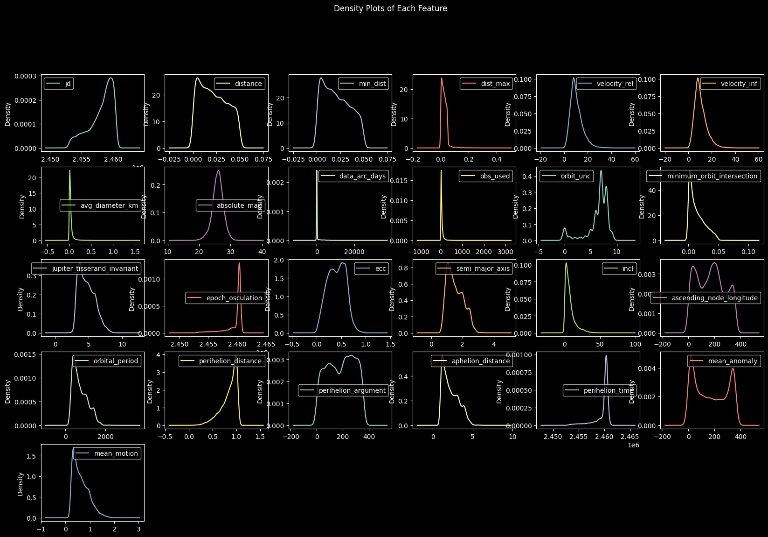


Fig. 1. Caption

## Model Development

The scaled features and label were split into training and testing sets. Because of a large class imbalance, Synthetic Minority Oversampling Technique (SMOTE) was used on the X and y training sets. This ensures the model does not become biased towards a majority class.

A Random Forest classifier is trained on the resampled training set and evaluates it on the test set, providing an unbiased estimate of its performance on unseen data. Then, A 5-fold cross validation on the entire dataset is performed.

# Results, Conclusion, and Future Work

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**