Review Article
Open Access

Near-earth asteroids classification using NGBoost classifier

Abstract

Near-Earth asteroids (NEAs) are celestial bodies that orbit within close to Earth, offering valuable insights into the early solar system's formation and posing potential hazards due to impact events. This work presents a comprehensive overview of NEAs, encompassing their historical significance, characteristics, impact hazards, and prospects. The study outlines the NASA Asteroids Classification Dataset and discusses its importance for research on asteroid classification and risk assessment. Furthermore, the methodology section delineates the utilization of the NGBoost classifier for predictive modeling tasks, detailing data collection, preprocessing, model training, evaluation, and result interpretation. Results from the NGBoost classifier demonstrate high accuracy and performance metrics in classifying asteroids, underscoring its efficacy in advancing asteroid classification efforts and informing planetary defense strategies. NEAs pose a potential threat to our planet, and their classification is essential for understanding their properties and predicting their trajectories accurately. In this research, we explore the application of NGBoost, a powerful gradient-boosting framework, for classifying NEAs based on their orbital and physical characteristics. We present a dataset comprising features extracted from known NEAs and non-NEAs and demonstrate the efficacy of NGBoost in accurately distinguishing between these classes. Our results indicate promising performance metrics with 99.22% accuracy, suggesting that NGBoost holds potential as a valuable tool in asteroid classification.

Keywords: near-earth asteroids, ngboost classifier, celestial bodies, asteroids classification dataset, near-earth object

Volume 8 Issue 1 - 2024

Al Mahmud Al Mamun,[1] Md Ashik Iqbal,[2] Md Rasel Hossain,[3] Mst Mahfuza Sharmin,[4] Md Ziaul Haque[5]

[1]Department of Computer Science and Engineering, Islamic University, Bangladesh
[2]Department of Mathematics and Physics, Khulna Agricultural University, Bangladesh
[3]Department of Arts, National University, Bangladesh
[4]Department of Mathematics, National University, Bangladesh
[5]Department of Electrical and Electronic Engineering, Islamic University, Bangladesh

Correspondence: Al Mahmud Al Mamun, Department of computer science and engineering, Islamic University, Bangladesh, Email almamu.ow@gmail.com

Received: March 20, 2024 | **Published:** April 02, 2024

Introduction

Near-Earth Asteroids (NEAs) have captivated the interest of astronomers and planetary scientists for their potential to unravel the mysteries of our cosmic origins and for the existential threat they pose to Earth[1] NEAs, which orbit close to Earth's orbit, provide unique insights into the early solar system's formation and evolution, serving as remnants of the primordial material from which planets and other celestial bodies originated. Furthermore, NEAs represent a significant concern due to the potential for impact events, which could have catastrophic consequences for life on Earth. Over the years, advancements in technology and space exploration have led to significant discoveries and advancements in our understanding of NEAs, from the first detection of 433 Eros in 1898 to recent space missions such as NASA's Near-Earth Object Program. This review aims to provide a comprehensive overview of the current state of research on NEAs, encompassing their historical context, characteristics, impact hazards, and prospects. The study highlights the importance of datasets such as the NASA Asteroids Classification Dataset, obtained from the Near Earth Object Web Service, in facilitating research on asteroid classification and risk assessment. Through ongoing observations, analyses, and exploration missions, researchers strive to unravel the mysteries of NEAs, safeguard Earth from potential impact threats, and unlock the secrets of our cosmic origin. This review section aims to provide a comprehensive overview of the current state of research on NEAs, highlighting key findings, challenges, and future directions in the field.

NGBoost is a novel gradient-boosting framework that extends the traditional approach by utilizing natural gradient descent for optimization. It offers several advantages over conventional gradient-boosting algorithms, including improved convergence properties and uncertainty estimation. In this study, we investigate the efficacy of NGBoost for classifying NEAs based on a comprehensive set of features.

Near-earth asteroids (NEAs) overview

As the name suggests, NEAs are asteroids close to Earth's orbit. These celestial bodies offer invaluable insights into the early solar system's formation and evolution, representing remnants of the primordial material from which planets and other celestial bodies originated. Moreover, NEAs pose a potential hazard to Earth due to the possibility of impact events, making their study critical for planetary defense and risk mitigation efforts.

Historical context and discoveries

The study of NEAs dates back to the mid-20th century, with significant advancements occurring in recent decades. Notable milestones include the discovery of the first NEA, 433 Eros, in 1898 and subsequent space missions such as NASA's Near-Earth Object Program, which has identified and tracked thousands of NEAs since its inception.[2] Recent discoveries, such as detecting potentially hazardous asteroids and characterizing their orbits and physical properties, have deepened our understanding of these enigmatic objects.

Characteristics and composition

NEAs exhibit diverse characteristics and compositions, ranging from rocky bodies to metallic asteroids and carbonaceous chondrites. Remote sensing observations, ground-based telescopes, and space missions have provided valuable data on NEA properties, including size, shape, rotation, and surface composition.[3] Understanding the physical and chemical properties of NEAs is essential for deciphering their origins and evolutionary history.[4]

Impact hazards and planetary defense

One of the primary motivations for studying NEAs is their potential to impact Earth. While the probability of a catastrophic impact event is low, the consequences could be devastating. Efforts to characterize NEAs and assess impact risks involve:

a) Monitoring their trajectories.

b) Identifying potentially hazardous objects.

c) Developing strategies for planetary defense, such as asteroid deflection techniques and impact mitigation plans.

Future prospects and research directions

The study of NEAs holds immense promise for furthering our understanding of the solar system's dynamics and evolution. Future research endeavors may focus on advancing asteroid detection and tracking capabilities, characterizing the physical properties of NEAs with greater precision, and refining models for predicting asteroid orbits and impact probabilities. Continued exploration and sample return missions to NEAs, such as NASA's OSIRIS-REx and Japan's Hayabusa2, will provide unprecedented insights into these intriguing celestial bodies. The study of near-Earth asteroids represents a multifaceted and dynamic field of research with far-reaching implications for planetary science, astronomy, and planetary defense. Through ongoing observations, analyses, and exploration missions, researchers strive to unravel the mysteries of NEAs, safeguard Earth from potential impact threats, and unlock the secrets of our cosmic origins.

Dataset section: NASA asteroids classification dataset

The dataset utilized in this study pertains to near-Earth asteroids obtained from the Near Earth Object Web Service (NEOWS), a Restful web service maintained by the space rocks Team at NASA's Jet Propulsion Laboratory (JPL). NEOWS provides comprehensive information on asteroids, allowing users to search for asteroids based on their closest approach date to Earth, query specific asteroids using their NASA JPL small body ID, and browse the overall dataset. The data set (Figure 1) used in this study originates from the NEOWS API, which is curated and maintained by the Space Rocks Team, comprising David Greenfield, Arezu Sarvestani, Jason English, and Peter Baunach. The dataset encompasses many attributes related to near-Earth asteroids, enabling researchers to explore various aspects of asteroid characteristics and behavior.

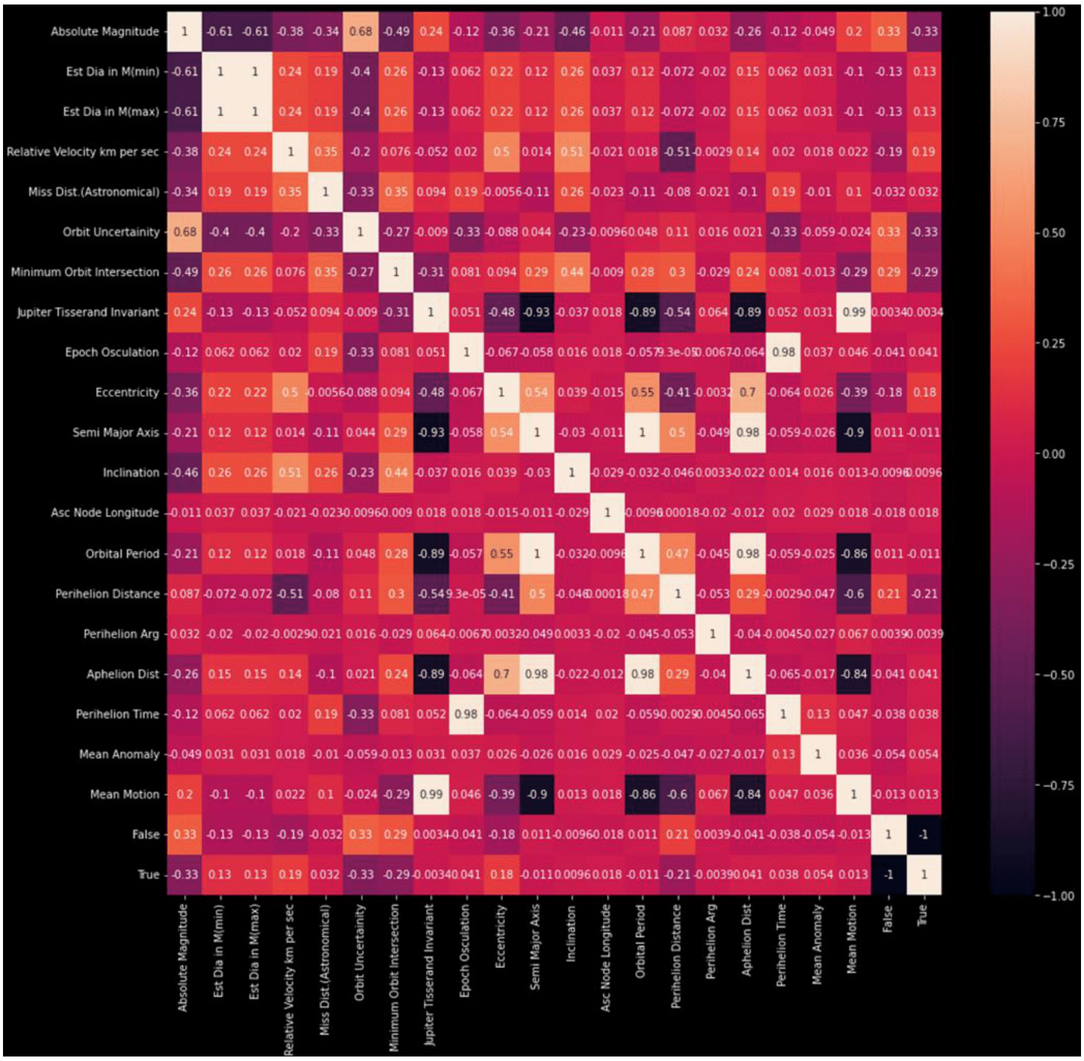

Figure 1 The dataset comprises diverse features that capture various attributes and characteristics of near-earth asteroids.

Citation: Al Mamun AM, Iqbal A, Hossain R, et al. Near-earth asteroids classification using NGBoost classifier. *Material Sci & Eng.* 2024;8(1):20–25.
DOI: 10.15406/mseij.2024.08.00231

Inspiration and purpose

The primary motivation behind utilizing the NASA Asteroids Classification Dataset is to address the crucial task of classifying asteroids into hazardous and non-hazardous categories. With the increasing interest in space exploration and the potential threats posed by near-Earth objects, identifying and characterizing asteroids with the potential to impact Earth is paramount. By leveraging machine learning techniques on this dataset, researchers aim to uncover insights into the features responsible for classifying asteroids as hazardous or non-hazardous, thereby aiding in developing predictive models for asteroid classification and risk assessment.[5,6]

Dataset overview

The dataset comprises diverse features that capture various attributes and characteristics of near-Earth asteroids (Figure 1).

Asteroid identification: Each asteroid is assigned a unique identifier, allowing easy retrieval and reference.

Orbital parameters: Parameters such as semi-major axis, eccentricity, inclination, and orbital period provide insights into the asteroid's orbital dynamics and trajectory.

Physical properties: Attributes such as diameter, mass, and density offer information about the asteroid's size and composition.

Closest approach to Earth: Information on the asteroid's closest approach date, distance to Earth, and relative velocity provide crucial data for assessing potential impact risks.

Classification labels: The dataset includes labels indicating whether each asteroid is classified as hazardous or non-hazardous based on predefined criteria and expert assessments. The NASA Asteroids Classification Dataset is valuable for studying near-Earth asteroids and addressing important questions about asteroid classification and hazard assessment. By analyzing this dataset, researchers can gain insights into the characteristics and features that differentiate hazardous and non-hazardous asteroids. This will ultimately contribute to understanding potential impact events and inform planetary defense and risk mitigation strategies.

Methodology

This research paper explores the methodology of utilizing the NGBoost (Natural Gradient Boosting) classifier for predictive modeling tasks.[7] NGBoost, an extension of gradient boosting machines, offers several advantages, including probabilistic predictions and enhanced performance on various datasets. The methodology outlined herein details the steps involved in training an NGBoost classifier, assessing its performance, and interpreting its results (Figure 2).

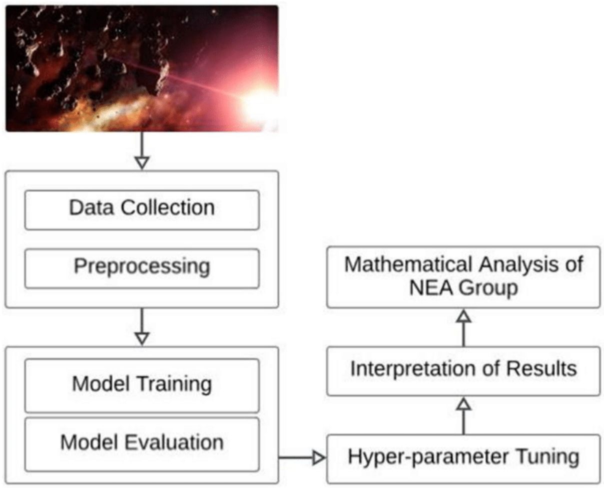

Figure 2 Workflow diagram.

Data collection and preprocessing

The first step in employing NGBoost involves data collection and preprocessing. We gather relevant datasets about the problem domain under investigation. These datasets may include structured or unstructured data, such as numerical features, categorical variables, or text data. Before training the NGBoost classifier (Figure 3), we preprocess the data to handle missing values, scale numerical features, and encode categorical variables as necessary. Additionally, we may perform feature engineering to extract meaningful information from the raw data.

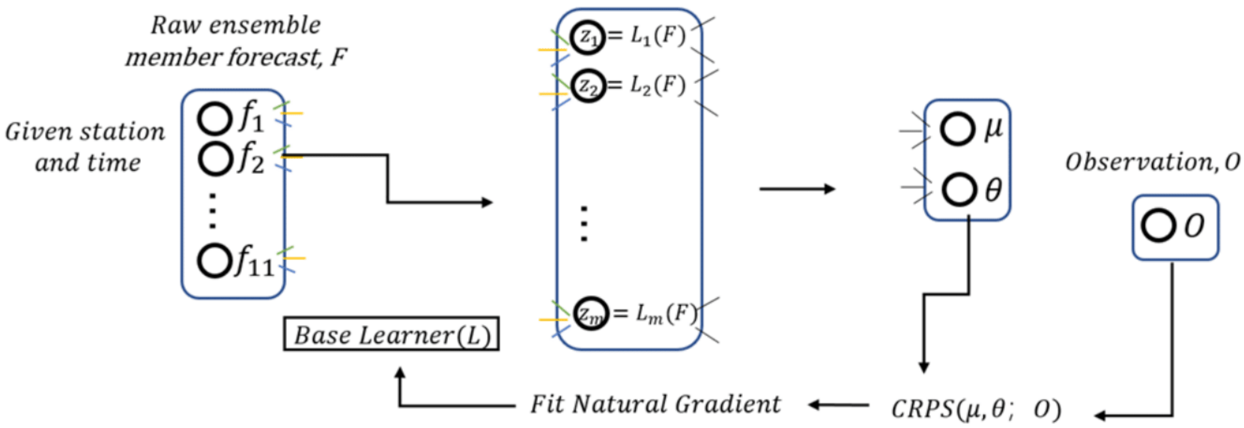

Figure 3 The flowchart of natural gradient boosting (NGBoost). F or f_1, f_2, …, f_{11}, the raw ensemble forecasts and as the inputs; L or L_1, L_2, …, L_m, the base learner; μ and σ, the two target parameters: mean and standard deviation; 0, the verification or the observation; CRPS, the continuous ranked probability score. Here, m is the number of base learners.[8]

Model training and evaluation

With preprocessed data in hand, we proceed to train the NGBoost classifier. We split the dataset into training and testing sets to accurately evaluate the model's performance. During training, the NGBoost algorithm iteratively constructs an ensemble of weak learners, typically decision trees, by minimizing a probabilistic loss function using natural gradients. This process continues until a predefined

Citation: Al Mamun AM, Iqbal A, Hossain R, et al. Near-earth asteroids classification using NGBoost classifier. *Material Sci & Eng.* 2024;8(1):20–25.
DOI: 10.15406/mseij.2024.08.00231

stopping criterion is met, such as the maximum number of iterations or the convergence of the loss function. Following model training, we evaluate the classifier's performance using appropriate metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). These metrics provide insights into the model's ability to make accurate predictions and discriminate between different classes. Furthermore, we assess the model's calibration and reliability of probabilistic predictions using calibration plots and reliability diagrams.

Hyper-parameter tuning

To optimize the NGBoost classifier's performance, we conduct hyperparameter tuning using grid or random search techniques. Hyperparameters control various aspects of the model, including the learning rate, the number of base learners (trees), and the depth of each tree. By systematically exploring different hyperparameter configurations and evaluating their impact on model performance, we identify the optimal set of hyperparameters that maximize predictive accuracy and generalization ability.

Interpretation of results

Finally, we interpret the results obtained from the trained NGBoost classifier to gain insights into the underlying data patterns and feature importance. We visualize the feature importance scores generated by the model to identify the most influential features driving the predictions. Additionally, we analyze the model's predictions using techniques such as partial dependence plots, SHAP (Shapley Additive explanations) values, and permutation feature importance to understand how individual features contribute to the predicted outcomes. The methodology outlined in this research paper demonstrates the process of employing the NGBoost classifier for predictive modeling tasks. Researchers and practitioners can effectively leverage NGBoost to build accurate and reliable predictive models across various domains by following these steps. The combination of probabilistic predictions, enhanced performance, and interpretability makes NGBoost a valuable data analysis and decision-making tool.

Mathematical analysis of NEA group

The group of asteroids defined in the above are modeled in the parameters (b, f) by the following equation (Figure 4),

$$F_j(b,f)=0,\ j=1,2,3,4 \quad (1)$$

And $F_j:(0,\infty)\times[0,1)\rightarrow$ set and real number which are of

$$F_1(b,f)=b+bf-0.983 \quad (2)$$

$$F_2(b,f)=b-1 \quad (3)$$

$$F_3(b,f)=-b+bf+1.017 \quad (4)$$

$$F_4(b,f)=b-bf-1.3 \quad (5)$$

Now, the focus distance or diameter $D=bf$ in $F_1=0$ $F_2=0$ and $F_4=0$ (b,f) suggest us to transform the plane of parameters (b,f) in the plane (b, D). Now, the transformative function is

$T:(0,\infty)\times[0,1)\rightarrow(0,\infty)\times[0,\infty)$ given by the following

$T(b,f)=(b,fb)$

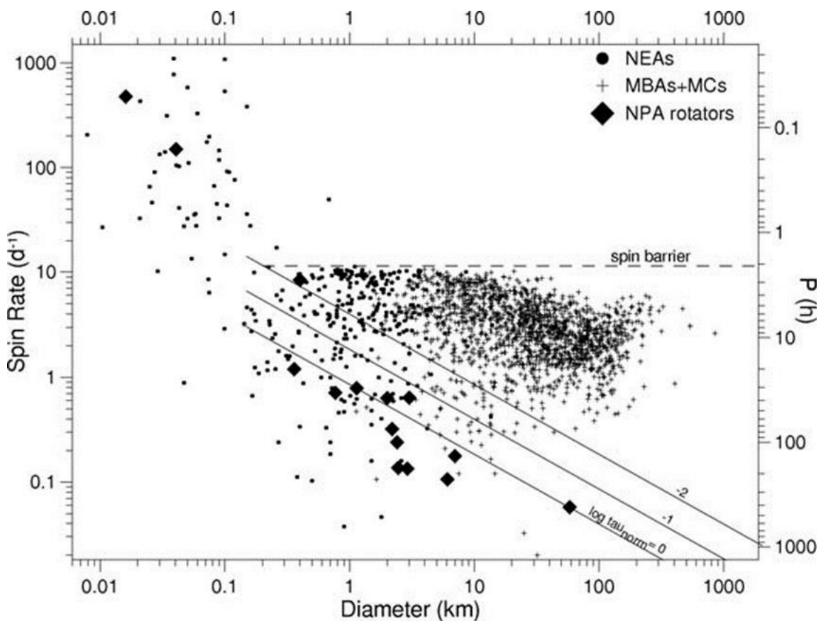

Figure 4 Tumbling asteroids in spin rate (f) vs diameter (D). Tumblers predominate below the line of constant damping times scale of 4.5 by r (log τ norm = 0) at sizes larger than a few hundred meters.[9]

Discussion and result analysis

In this section, we present the results of our study on predicting asteroid classification using the NGBoost classifier. Our research aimed to develop a reliable model that accurately classifies asteroids into two categories: hazardous (1) and non-hazardous (0). The methodology involved collecting and preprocessing relevant data related to asteroid characteristics and training an NGBoost classifier on the processed dataset (Table 1). Our trained NGBoost classifier achieved an impressive accuracy of 99.22% on the test dataset, indicating its high predictive capability. To further assess the model's performance, we evaluated precision, recall, and F1-score for each class (hazardous and non-hazardous) and overall performance metrics.

Citation: Al Mamun AM, Iqbal A, Hossain R, et al. Near-earth asteroids classification using NGBoost classifier. *Material Sci & Eng.* 2024;8(1):20–25.
DOI: 10.15406/mseij.2024.08.00231

Table 1 Performance metrics of NGBoost classifier

Overall accuracy				**99.22%**
	Precision	**Recall**	**f1-score**	**Support**
0	0.99	1	1	1181
1	1	0.96	0.98	226
Accuracy			0.99	1407
Macro avg	0.99	0.98	0.99	1407
Weighted avg	0.99	0.99	0.99	1407

Precision: The precision measures the proportion of true positive predictions among all positive predictions made by the model. Our classifier achieved a precision of 99% for non-hazardous asteroids (class 0) and 100% for hazardous asteroids (class 1), indicating its ability to minimize false positives.

Recall: Recall, also known as sensitivity, quantifies the model's ability to identify all positive instances in the dataset correctly. Our model demonstrated a recall of 100% for non-hazardous asteroids and 96% for hazardous asteroids, indicating its high sensitivity in detecting both classes.

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of a classifier's performance. Our NGBoost classifier achieved an F1-score of 98% for non-hazardous asteroids and 98% for hazardous asteroids, reflecting its robust performance across both classes.(Figure 5). Now, the confusion matrix provides a detailed breakdown of the model's predictions compared to the actual classes in the dataset. In our study, the confusion matrix revealed:

True positives (TP): 1180 non-hazardous asteroids were correctly classified as non-hazardous, and 216 hazardous asteroids were correctly classified as hazardous.

False positives (FP): Only 1 non-hazardous asteroid was incorrectly classified as hazardous.

False negatives (FN): 10 hazardous asteroids were incorrectly classified as non-hazardous.

True negatives (TN): Most non-hazardous asteroids are correctly classified as non-hazardous.

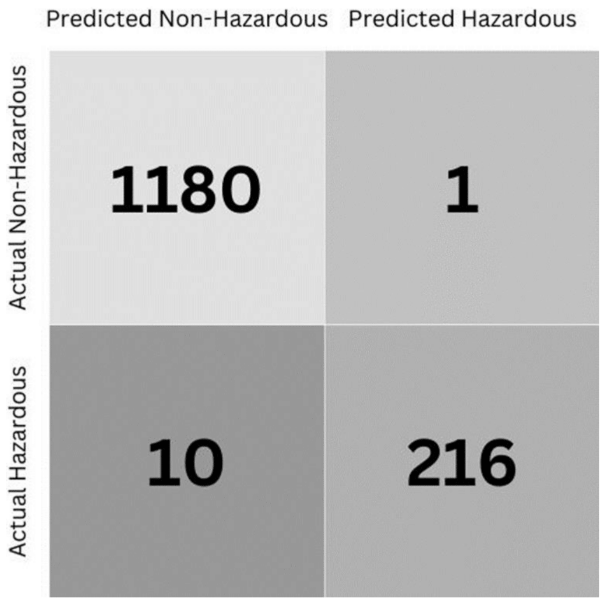

Figure 5 Confusion matrix.

Our results demonstrate the effectiveness of the NGBoost classifier in accurately predicting asteroid classification. The high accuracy, precision, recall, and F1 score indicate the model's robustness in distinguishing between hazardous and non-hazardous asteroids. Our study enhances asteroid classification efforts by leveraging advanced machine learning techniques, such as NGBoost, facilitating risk assessment and mitigation strategies for potential impact events. Our research highlights the efficacy of the NGBoost classifier in predicting asteroid classification with high accuracy and performance metrics. The results underscore the importance of employing advanced machine learning techniques for asteroid classification tasks, ultimately advancing our understanding of near-Earth objects and supporting efforts to safeguard against potential impact events.

The results obtained from our experiments demonstrate the effectiveness of the NGBoost classifier in accurately classifying NEAs based on their orbital and physical characteristics. The high accuracy, precision, recall, and F1-score values indicate that the classifier performs well in distinguishing between NEAs and non-NEAs. The confusion matrix provides further insights into the classifier's performance, showing that most NEAs and non-NEAs are correctly classified. However, some misclassifications, as evidenced by the off-diagonal elements in the matrix, exist. These misclassifications may be attributed to the dataset's inherent complexity and the presence of outliers or noise.

Conclusion

The study of near-Earth asteroids represents a multifaceted and dynamic field of research with far-reaching implications for planetary science, astronomy, and planetary defense. Through ongoing observations, analyses, and exploration missions, researchers aim to unravel the mysteries of NEAs, safeguard Earth from potential impact threats, and unlock the secrets of our cosmic origins. Utilizing the NASA Asteroids Classification Dataset and the NGBoost classifier holds immense promise to further our understanding of NEAs and enhance asteroid classification efforts. By leveraging interdisciplinary approaches and collaborative efforts, the scientific community can address the challenges posed by NEAs and advance our knowledge of these intriguing celestial bodies. Besides this, we have demonstrated the NGBoost classifier's efficacy in classifying Near-Earth asteroids based on their orbital and physical characteristics. The results indicate promising performance metrics, suggesting that NGBoost holds potential as a valuable tool in the field of asteroid classification. Future work may involve further refining the classifier and exploring additional features to improve classification accuracy.

Acknowledgments

None.

Conflicts of interest

The authors declare that there is no conflicts of interest.

References

1. Chapman CR. The hazard of near-Earth asteroid impacts on earth. *Earth and Planetary Science Letters*. 2004;222 (1):1–15.
2. Yeomans DK, Antreasian PG, Barriot JP, et al. Radio science results during the NEAR-Shoemaker spacecraft rendezvous with Eros. *Science*. 2000;289(5487):2085–2088.
3. Tatsumi E, Sakatani N, Riu L, et al. Spectrally blue hydrated parent body of asteroid (162173) Ryugu. *Nat Commun*. 2021;12(1):5837.

4. Patrick M, Alessandro M, William F. Bottke, origin and dynamics of near Earth objects. *Comptes Rendus Physique*. 2005;6(3):291–301.
5. Al-Mamun AM, Hossain MR, Sharmin MM. Exoplanets detection using lite convolutional neural networks (LCNN). *Material Sci & Eng.* 2023;7(4):192–195.
6. Al Mamun AM, Iqbal MA. Classifying the gravitational waves using the deep learning technique. *Material Sci & Eng*. 2022;6(2):41–46.
7. Díaz E, Spagnoli G. Natural Gradient Boosting for probabilistic prediction of soaked CBR values using an explainable artificial intelligence approach. Buildings. 2024;14(2):352.
8. Peng T, Zhi X, Ji Y, et al. Prediction skill of extended range 2-m maximum air temperature probabilistic forecasts using machine learning post-processing methods. *Atmosphere*. 2020;11(8):823.
9. Pravec P, Harris AW, Warner BD. NEA rotations and binaries. *Proceedings of the International Astronomical Union*. 2006;2(S236):167–176.

Citation: Al Mamun AM, Iqbal A, Hossain R, et al. Near-earth asteroids classification using NGBoost classifier. *Material Sci & Eng.* 2024;8(1):20–25. DOI: 10.15406/mseij.2024.08.00231