**A Project on**

**NASA: Asteroids Classification**

***Submitted in partial fulfillment of the requirement for the award of the degree of***

Masters Of Computer Application

****

**Under The Supervision of**

**Dr. Mala Saraswat**

**Associate Professor**

**Submitted By**

**Hensika Bishnoi (E23MCAG0004)**

**Eeva Mehra (E23MCAG0049)**

**Aakriti Jain (E23MCAG0131)**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING AND TECHNOLOGY**

**BENNETT UNIVERSITY, GREATER NOIDA, U.P**

# School of Computer Science Engineering andTechnology - Bennett University

**CANDIDATE’S DECLARATION**

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **“NASA: Asteroids Classification”** in partial fulfillment of the requirements for the award of the **Masters of Computer Application** submitted in the School of Computing Science and Engineering of **Bennett University, Greater Noida**, is an original work carried out during the period of month, Year to Month and Year, under the supervision of **Dr. Mala Saraswat**, Associate Professor School of Computer Science Engineering and Technology, Bennett University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Names – Hensika Bishnoi, Eeva Mehra, Aakriti Jain

**CERTIFICATE**

The Project ……… of ………………

has been held on and his/her work is recommended for the award of Master of Computer Applications.

**Signature of Examiner(s) Signature of Supervisor(s)**

**Signature of Program Chair Signature of Dean**

Date: May, 2024 Place: Greater Noida

# Abstract

This project delves into the classification of asteroids through machine learning methods, utilizing NASA's extensive dataset. The main aim is to develop models capable of accurately classifying asteroids based on various characteristics like size, composition, and orbital parameters. By employing machine learning algorithms, the study seeks to uncover patterns within the asteroid data to enable effective classification. Challenges such as data scarcity, noise, and class imbalance are addressed through preprocessing and modelling techniques like data augmentation, transfer learning, and ensemble methods. Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the performance of each model, with a focus on identifying the most suitable approach for asteroid classification. Insights derived from this project contribute to scientific understanding of asteroids and have practical implications for planetary defense and space exploration. Accurate classification enables better risk assessment and the development of mitigation strategies. Additionally, the application of machine learning in asteroid classification highlights the potential for advancements in space science and technology, facilitating improved monitoring and comprehension of celestial bodies within our solar system.

**Keywords - nasa, asteroids, machine learning, hazardous, classification**

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **Title** |  | **Page No.** |
| **Abstract** |  | **I** |
| **List of Table** |  | **II** |
| **List of figures** |  | **III** |
| **Chapter 1** | **Introduction**   * 1. Motivation   2. Problem statement   3. Aim and objectives of the study | **1** |
| **Chapter 2** | **Literature Survey**   * 1. A review of relevant literature and existing work related to the project (with table). |  |
| **Chapter 3** | **Project Design**   * 1. Hardware/ Software Requirement   2. Proposed Methodology   3. Details on data collection methods and tools   4. Procedures and techniques used in the project |  |
| **Chapter 4** | **Result and Discussion**   * 1. Presentation of the findings or results of the project   2. Use of tables, charts, graphs, or other visuals for clarity. Explanation of result. |  |
| **Chapter 5** | **Conclusion and Future Work**  **REFERENCES** |  |

**Chapter 1: Introduction**

This study's ideas arise from how crucial it is to learn about asteroids and how they may impact on Earth. By accurately classifying asteroids as hazardous or non-hazardous, we can better assess the risk they act and contribute to space science and planetary defense efforts. The classification of asteroids is central to the problem statement; non-hazardous asteroids are essential for understanding the structure of our solar system, while hazardous asteroids could pose a threat to Earth if their orbits collide with ours. Creating an effective classifying model to forecast the degrees of asteroid hazard is the goal of this research. A number of goals are specified in order to accomplish this goal: feature engineering, data exploration and preprocessing, model evaluation, training and selection, interpretation, and conclusion. Through these objectives, we seek to enhance our understanding of asteroids and their hazard levels, ultimately contributing to planetary defense initiatives.

**Objectives of the study**

1. Develop machine learning models for accurately classifying asteroids based on their characteristics like size, composition, and orbital parameters.
2. Gain insights into asteroid properties and behaviours through analysis of available NASA asteroid data.
3. Assess risks posed by hazardous asteroids and identify strategies for risk mitigation.
4. Contribute to scientific understanding of asteroids, including their origins, evolution, and potential impact on Earth.
5. Explore the application of machine learning techniques in asteroid classification and monitoring to drive advancements in space science and technology.
6. Enhance preparedness and response capabilities for potential asteroid impacts, safeguarding human lives and infrastructure on Earth.

**Chapter 2: Literature Survey**

Through the years, there have been many developments in the field of classification of Near-Earth Objects. The authors provide diverse perspectives on asteroid classification and hazard assessment using machine learning techniques. Klimczak et al., 2021 [1] emphasizes the classification of asteroids by leveraging machine learning and spectral features to replicate the Bus-DeMeo taxonomy and optimize future surveys. Malakouti, 2023 [2] focuses on hazard classification, employing algorithms like Random Forest to identify hazardous asteroids through meticulous analysis of NASA data. Carruba et al., 2019 [3] and Carruba et al., 2020 [4] delve into machine learning algorithms for asteroid family identification and compare classification algorithms, respectively. Paper [3] showcases superior accuracy over traditional methods, while Ramakrishnan, n.d. [5] evaluates the performance of algorithms, with Random Forest and XGBoost emerging as top performers. The distribution and classification of basaltic asteroids are explored in Mansour et al., 2020 [6], while a comparative analysis of ML algorithms for asteroid taxonomic classification is presented in Klimczak et al., 2022 [7]. Advancements in Quantum Machine Learning for asteroid hazard prediction are discussed in Bhavsar et al., 2023 [8], whereas deep convolutional neural networks are employed for asteroid classification in Bacu et al., 2023 [9]. A novel approach utilizing machine learning for asteroid classification based on meteorite spectral data is proposed by Dyar et al., 2023 [10]. Papers Machado et al., 2022 [11] and Bahel et al., 2021 [12] propose quantitative approaches for identifying hazardous asteroids, employing clustering tools and supervised machine learning methods, respectively. These diverse studies have altogether contributed to the growing research on asteroid classification and hazard assessment, showcasing the potential of machine learning in advancing our understanding of celestial bodies.

**Chapter 3: Project Design**

**Dataset Description**

The dataset contains information about asteroids and labels each asteroid as either hazardous or non-hazardous. Some of the prominent features are -

* The dataset is collected from the **NASA API**.
* It’s an aggregation of **27 PDS data sets** with tabulations and compilations of asteroid data on CD-WO.
* **This dataset** contains **4687 data instances (rows)** and **40 features (columns)**. Each row represents information about an asteroid, and the columns provide various attributes related to these celestial objects.
* The dataset includes various attributes related to asteroids, such as -
  + **Orbit characteristics**: Information about the asteroid’s orbit, including its semi-major axis, eccentricity, inclination, and orbital period.
  + **Physical properties**: Details like diameter, albedo, and absolute magnitude.
  + **Close approach data**: Date when the asteroid comes closest to Earth.
  + **Hazard classification**: Whether the asteroid is hazardous or not.

**Methodology**

1. *Data Collection and Exploration*
   1. Gather data from the NASA Asteroids Classification dataset.
   2. Explore the dataset to understand its features, distributions, and characteristics.
2. *Data Preprocessing*
   1. Handle any missing values, outliers in the dataset.
   2. Perform data cleaning and normalization the data
3. *Feature Engineering*
   1. Extract relevant features from the dataset.
   2. Select features from the dataset that are informative for asteroid classification.
4. *Model Selection and Training*
   1. Choose appropriate machine learning algorithms for classification (e.g., Decision Trees, Random Forest, Support Vector Machines).
   2. Train multiple models using labeled data to predict asteroid hazard levels.
5. *Model Evaluation*
   1. Assess the performance of each model using evaluation metrics such as accuracy, precision, recall, and F1 score.
   2. Validate the models using cross-validation techniques to ensure generalization to unseen data.

For the purpose of achieving the research goals, a number of classifiers were trained and evaluated as part of the process. To expedite the training and evaluation technique, a function known as "classifiers" was initially implemented. We chose seven classifiers for testing: Random Forest, K-Nearest Neighbors, Decision Tree, Logistic Regression, SVM, Gaussian Naive Bayes, and Light GBM Classifier. The training dataset was used to train each classifier, and the test dataset was used to evaluate its performance. Each classifier's performance was evaluated using performance criteria like accuracy, precision, recall, and F1 score. In addition, confusion matrices were produced for the training and test sets, offering visual representations of the classification results.

In order to help with model selection and further the goals of the study, the accuracy, precision, and recall scores that were obtained were noted and methodically compared. The NASA Asteroids Classification dataset was analyzed, and seven classifiers with different features were investigated. With a focus on feature independence, Gaussian Naive Bayes (GaussianNB) is known for its effectiveness and simplicity. Although very good at identifying the best hyperplanes, Support Vector Machine (SVM) can be computationally demanding. Decision Trees can be easily interpreted, but often face the issue of overfitting. In addition to producing interpretable coefficients, Logistic Regression works well with linear bounds. K-Nearest Neighbors (KNN) is comparatively easier to choose, but at the same time can become slow when dealing with big datasets. Random Forest handles high-dimensional data and minimizes overfitting effectively. And lastly, Light GBM excels in efficiency and accuracy, particularly for large-scale datasets.

**Chapter 4: Results and Discussion**

A thorough examination of different machine learning models was performed to evaluate their effectiveness in classifying asteroids. A wide range of algorithms were implemented in this study, such as Decision Tree, Random Forest, LightGBM, K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GaussianNB), Support Vector Machine (SVM), and Logistic Regression models. The evaluation metrics of each algorithm is displayed in Table 1.

Table 1. Evaluation metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S. No.** | **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| 1. | Decision Tree | 0.992537 | 0.951724 | 1.000000 | 0.979021 |
| 2. | Random Forest | 0.992537 | 0.951724 | 1.000000 | 0.975265 |
| 3. | LightGBM | 0.991471 | 0.951724 | 0.992806 | 0.971831 |
| 4. | KNN | 0.846482 | 0.006897 | 1.000000 | 0.013699 |
| 5. | Gaussian NB | 0.845416 | 0.000000 | 0.000000 | 0.000000 |
| 6. | SVM | 0.845416 | 0.000000 | 0.000000 | 0.000000 |
| 7. | Logistic Regression | 0.845416 | 0.000000 | 0.000000 | 0.000000 |

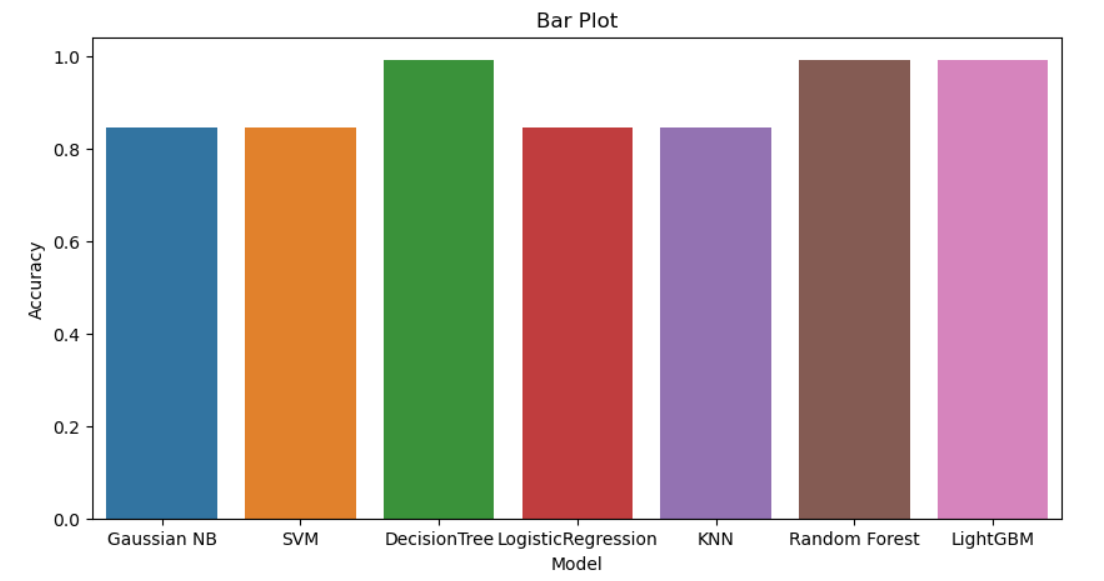
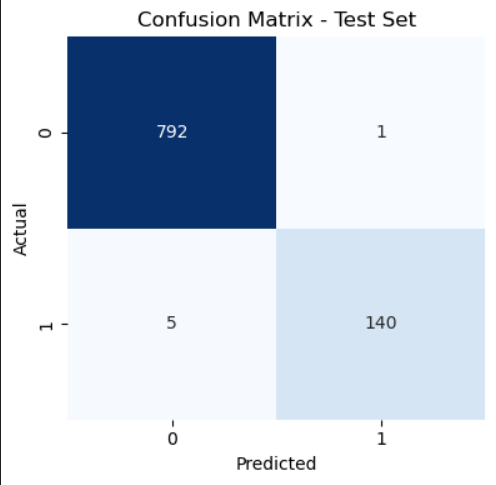
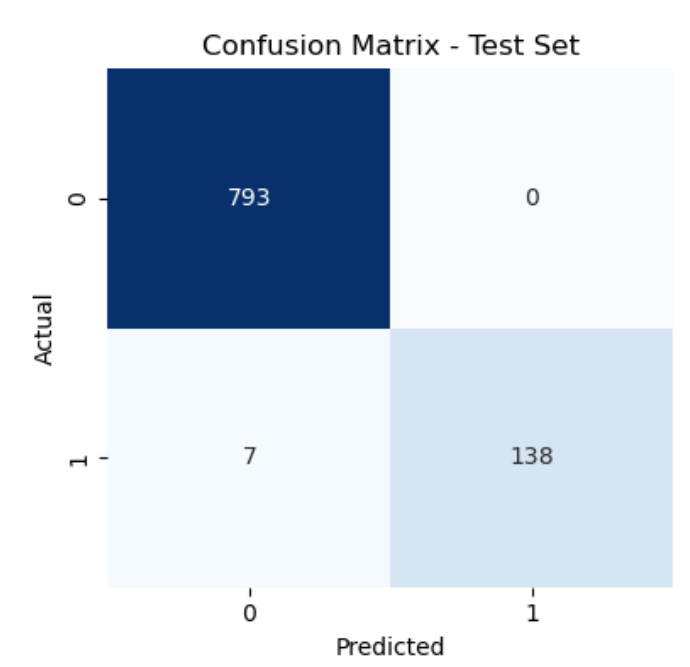


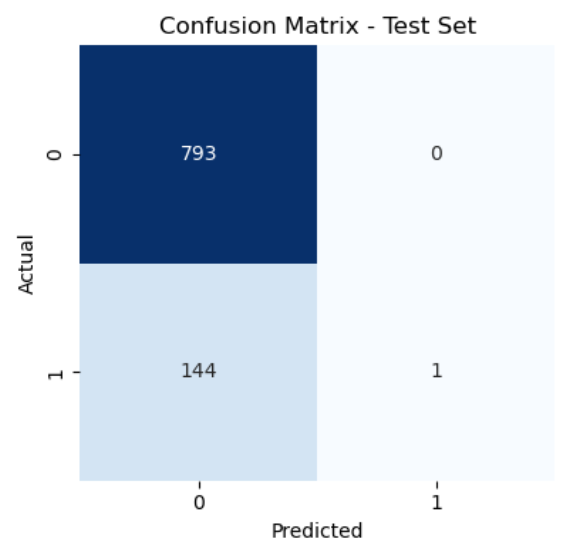
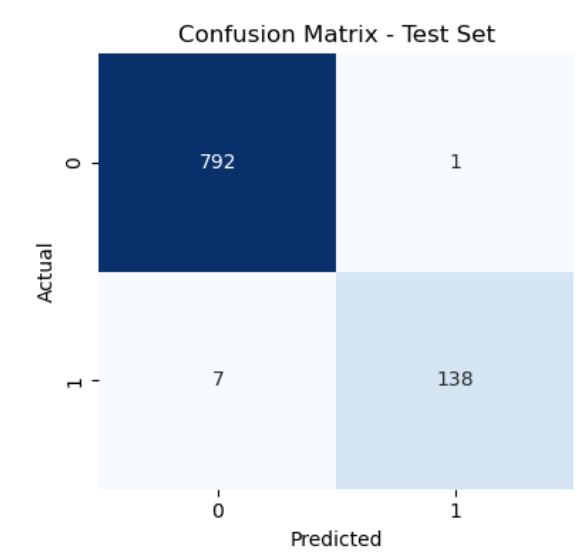
Fig. 1 Accuracy bar plot

The evaluation metrics display the outcome of our analysis which indicate that the Decision Tree and Random Forest models were superior to the other models in terms of accuracy (see Fig. 1), where the score of these models show that they excel at recognizing intricate patterns and relationships in the data making them well-suited for classifying asteroids. The LightGBM closely trailing behind, achieved an accuracy of 0.991471 and showed strong performance. This advanced machine learning technique, known as gradient boosting, is specifically designed to manage large datasets effectively, and it has been found to be a significant improvement in our model comparison. While the KNN model (k = 10) had a lower accuracy, it is still appealing for certain uses where finding all positive examples is the main goal.

The confusion matrices for these models are displayed in Fig. 2.



(a) Decision Tree (b) Random Forest



(c) Light GBM (d) KNN

Fig. 2 Confusion matrices

The Gaussian NB, SVM, and Logistic Regression models, however, demonstrated the lowest accuracy scores, ranging from 0.845416 to 0.846482. While these models may not be optimal for asteroid classification with the current dataset and evaluation metrics, they still hold value for other applications and can serve as useful baselines for comparison.

**Chapter 5: Conclusion and Future Work**

In summary, this study has conducted a thorough assessment of diverse machine learning models for asteroid classification, emphasizing the strengths and weaknesses of each approach. The Decision Tree and Random Forest models have emerged as the top performers, showcasing their potential as dependable tools for asteroid classification tasks. However, it is vital to weigh the trade-offs between different models when selecting the most suitable one for a specific application. Considerations such as interpretability, computational efficiency, and resilience to noise and outliers should also be taken into consideration.

This study adds to the growing knowledge base in asteroid classification and can be used as a basis for future research. Further research could include examining additional models, refining hyperparameters, and integrating advanced feature engineering techniques to enhance classification accuracy. By continually improving and broadening our comprehension of machine learning models in asteroid classification, we can ultimately improve our capacity to analyze and understand the extensive data produced by contemporary astronomical observations.

**REFERENCES**

[1] Klimczak, H., Kotłowski, W., Oszkiewicz, D., DeMeo, F., Kryszczyńska, A., Wilawer, E., & Carry, B. (2021). Predicting Asteroid Types: Importance of Individual and Combined Features. *Frontiers in Astronomy and Space Sciences*, *8*. <https://doi.org/10.3389/fspas.2021.767885>

[2] Malakouti, S. M. (2023). Discriminate primary gammas (signal) from the images of hadronic showers by cosmic rays in the upper atmosphere (background) with machine learning. *Physica Scripta*, *98*(4). <https://doi.org/10.1088/1402-4896/acc1b2>

[3] Carruba, V., Aljbaae, S., & Lucchini, A. (2019). Machine-learning identification of asteroid groups. *Monthly Notices of the Royal Astronomical Society*, *1*, 1377–1386. <https://doi.org/10.1093/mnras/stz1795>

[4] Carruba, V., Aljbaae, S., Domingos, R. C., Lucchini, A., & Furlaneto, P. (2020). Machine learning classification of new asteroid families members. Monthly Notices of the Royal Astronomical Society, 496(1), 540–549. <https://doi.org/10.1093/mnras/staa1463>

[5] Ramakrishnan, A. (n.d.). Hazardous Asteroid Classification with Machine Learning using Physical and Orbital Asteroid Properties.

[6] Mansour, J. A., Popescu, M., De León, J., & Licandro, J. (2020). Distribution and spectrophotometric classification of basaltic asteroids. Monthly Notices of the Royal Astronomical Society, 491(4), 5966–5979. <https://doi.org/10.1093/mnras/stz3284>

[7] Klimczak, H., Oszkiewicz, D., Carry, B., Penttilä, A., Kotlowski, W., Kryszczyaska, A., & Wilawer, E. (2022). Comparison of machine learning algorithms used to classify the asteroids observed by all-sky surveys. Astronomy and Astrophysics, 667. <https://doi.org/10.1051/0004-6361/202243889>

[8] Bhavsar, R., Jadav, N. K., Bodkhe, U., Gupta, R., Tanwar, S., Sharma, G., Bokoro, P. N., & Sharma, R. (2023). Classification of Potentially Hazardous Asteroids Using Supervised Quantum Machine Learning. IEEE Access, 11, 75829–75848. <https://doi.org/10.1109/ACCESS.2023.3297498>

[9] Bacu, V., Nandra, C., Sabou, A., Stefanut, T., & Gorgan, D. (2023). Assessment of Asteroid Classification Using Deep Convolutional Neural Networks. Aerospace, 10(9). <https://doi.org/10.3390/aerospace10090752>

[10] M. Darby Dyar, Sydney M. Wallace, Thomas H. Burbine, Daniel R. Sheldon, A machine learning classification of meteorite spectra applied to understanding asteroids, Icarus, Volume 406, 2023, 115718, ISSN 0019-1035, <https://doi.org/10.1016/j.icarus.2023.115718>.

[11] Machado, J.A.T., Mehdipour, S.H. Multidimensional Analysis of Near-Earth Asteroids. *SN COMPUT. SCI.* **3**, 207 (2022). <https://doi.org/10.1007/s42979-022-01103-2>

[12] V. Bahel, P. Bhongade, J. Sharma, S. Shukla and M. Gaikwad, "Supervised Classification for Analysis and Detection of Potentially Hazardous Asteroid," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 2021, pp. 1-4, doi: 10.1109/ICCICA52458.2021.9697222.