2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)

Classification and Comparative Analysis of Hazardous Asteroids using Machine Learning

Aarav Srivastava
School of Computer Engineering
KIIT Deemed to be University
Bhubaneshwar, Odisha, India
srivastava_aarav@outlook.com

Mahendra Kumar Gourisaria
School of Computer Engineering
KIIT Deemed to be University
Bhubaneshwar, Odisha, India
mkgourisaria2010@gmail.com

Anjishnu Saw
School of Computer Engineering
KIIT Deemed to be University
Bhubaneshwar, Odisha, India
23051895@kiit.ac.in

Dayal Kumar Behara
School of Computer Engineering
KIIT Deemed to be University
Bhubaneshwar, Odisha, India
dayalbehera@gmail.com

Sonal Jain
School of Economics and Commerce
KIIT Deemed to be University
Bhubaneshwar, Odisha, India
sonalhzbjain@gmail.com

Tanvir Habib Sardar
CSE, School of Engineering
Dayanand Sagar University
Bangalore, Karnataka, India
tanvir-cse@dsu.edu.in

Abstract— **Asteroids are small, rocky celestial objects that continuously orbit the Sun. They can be hazardous if classified as Near-Earth Objects (NEOs). Machine learning models can assist in predicting the nature of asteroids and classify them as potentially hazardous or not. The aim of this study is to optimize the identification process of Potentially Hazardous Asteroids (PHAs) and Near-Earth Asteroids (NEAs). In this work, we have implemented several classification and regression models and validated their performance. We implemented Random Forest, Naive Bayes, K-Nearest Neighbors and Logistic Regression classifiers. For regression tasks, we implemented Linear Regression, Ridge Regression and Lasso Regression. We used performance metrics such as Accuracy, Precision, Recall, and F1 Score, and regression evaluation metrics including RMSE and R^2 Score, to assess model performance. Results show that the best performance was achieved by the Random Forest classifier, with an accuracy of 99.99%.**

Keywords—Potentially Hazardous Asteroids (PHA), Near Earth Asteroids (NEA), Performance Metrics, Predictive Modeling, Classification.

I. INTRODUCTION

Asteroids are rocky fragments without an atmosphere, remaining from the early formation of our solar system about 4.6 billion years ago [1]. They are also regarded as minor planets since they orbit the Sun just like Earth and other planets in the solar system, although they are much smaller in size. Most asteroids can be found within the main asteroid belt located between Mars and Jupiter. Asteroids are regarded as one of the most critical and influential topics in space research as it comprises of large number of objects - stars, planets, moons, asteroid and it is highly important to predict the potentially hazardous nature of asteroids. The field of research associated with the analysis and prediction of the hazardous nature of asteroids has been tremendously progressive, thanks to consistent improvements in technology and scientific understanding. Since there exists various types and classes of asteroids, the process of target selection is a key factor of designing an asteroid mission. It is advised to take technological viability and scientific interest into consideration [2].

The overall process of identifying and capturing key insights of asteroids that could pose a threat to Earth is known as asteroid hazard prediction [3]. The main goal of our model is to detect and determine asteroids capable of severely impacting Earth causing significant and catastrophic damage. The implementation of machine learning techniques has proven advantageous in the domain of asteroid hazard prediction in recent times. Majority of individuals have a strong belief that an asteroid is classified as dangerous or threatening only on the basis of its size when it is about to collide with Earth. However, factors other than the size of the asteroid also contribute in determining the precarious impact the asteroid can have, such as mass of the asteroid, the velocity at which it is expected to the strike the surface of the Earth, NEO flag, PHA flag etc. Machine learning encounters several obstacles in asteroid classification such as unbalanced datasets, sparse data and trouble in extrapolation of new data. Further challenges are presented by the subjective nature of visual inspection in existing classification algorithms and small spectral changes within the asteroid classes. It is clearly a complex task to accurately predict the true nature of any asteroid by merely analyzing the size of the asteroid or the amount of energy it radiates when it strikes the surface of the Earth. Hence, machine learning models are needed for the identification, classification and prediction of NEO and PHA flag for each and every asteroid. NEO (Near-Earth Object) is defined as an object (here, asteroid) whose orbit is in within 1.3 astronomical units of the sun. PHA (Potentially Hazardous Asteroid), which in turn is also a near Earth object is at least 140 meters in size and whose orbit approaches the Earth's object within 0.05 AU (74,80,000 kilometers).

The use of machine learning models significantly assists in the prediction of hazardous nature of asteroids. The various aspects of Data Science and Machine Learning utilized were: Data Exploration, Data Cleaning, Data Preprocessing, Model Developing, Model Engineering. Recently, various kinds of machine learning algorithms are being developed and optimized leading in greater abundance and availability for solving classification problems. Using an existing dataset, it is possible to utilize these algorithms and train the models accordingly to efficiently predict if a new set of data belongs to a particular group [4]. Careful analysis of different kinds of algorithms allows researchers to accurately predict which method performs the best and proficiently addresses the problem statement. ML algorithms can also be used in different other fields like virus detection [5], cancer prediction [6], osteoporosis prediction [7] etc. It is critical for us to save the planet by improving our skills and knowledge in the field of research and machine learning and putting efforts into successful detection of hazardous asteroids and its mitigation techniques.

979-8-3315-1308-5/25/$31.00 ©2025 IEEE

The remainder of the paper is organized and structured as per the following format: Section II. Related Work, covering significant publications on the subject by several other researchers; Section III. tells about the Materials and Methods, providing details on the classifiers and strategies employed along with the technology used; Section IV covers Result and Discussion, contrasting the performance of various classifiers to determine the most effective model; and Section V. Conclusion and Future Work, summarizing findings and proposing future research directions followed by References.

II. Related Work

Machine learning models are becoming more and more popular and useful in predicting the hazardous nature of asteroids which is nothing but a classical binary classification problem statement. This section discusses about the different kinds of techniques implemented and algorithms utilized by other researchers in training their models and constructs a comparative analysis of their performance and accuracy.

N. P. Surekha et al. [8] (2024) have extracted the same dataset from Kaggle and trained the model on a mix of algorithms including ensemble, regression and boosting algorithms - Logistic Regression, Random Forest and Light Gradient Boosting. The most accurate results were obtained using LightGBM with importance. M. Sam Kennaya et al. [9] (2023) extracted the dataset right from the source i.e. Jet Propulsion Laboratory. The data was cleaned and processed further by the means of label encoding to convert categorical values into numeric values. Features were selected on the basis of their feature importance score through the Random Forest classifier model and reduced the number from 45 to 15 for better and efficient training and testing. The model was trained on an Artificial Neural Network by the means of oversampling techniques such as SMOTE. This helped the authors in obtaining an accuracy of 99.86% using borderline SMOTE.

V. Bahel et al. [10] (2021) have utilized the asteroid dataset sourced from "Small-Body Database" which is also provided by the Jet Propulsion Laboratory. The authors split the data in the ratio of 1:2 (testing to training). The algorithms that were used by them were Logistic Regression, Random Forest, K Nearest Neighbor and Decision Trees, obtaining the best performance score through Random Forest i.e. an accuracy of 99.99%. RKMT N. Ranaweera et al. [11] (2022) proposed their work by utilizing a dataset from JPL Horizon. The data was highly skewed and imbalance as it had 10,32,040 asteroids classified as non-hazardous but only 2,059 labeled as hazardous. To tackle this problem, the authors prepared and developed a wrapper class from PyTorch and constructed a PHAC (Potentially Hazardous Asteroid Classifier) through the dynamics of domain knowledge and deep learning techniques. The authors were able to develop a robust model using KNN algorithm with an accuracy of 91%.

A. Sandeep et al. [12] (2024) self-procured a dataset of asteroids consisting of columns like Designation, Discovery Date, H (magnetic field intensity), MOID (minimum orbit intersection distance), q (perihelion distance in astronomical units) and orbit class of the asteroid. Duplicate and null values in the dataset were handled and removed and the model was further split and trained on several machine learning algorithms, obtaining the most accurate performance score through XGBoost algorithm i.e. 100%. Lastly, L. Almousa et al. [13] (2025) utilized the dataset from Minor Planet Bodies, which was then uploaded to Kaggle. The dataset contained 1,02,510 instances. The columns and features which contained null values were simply dropped and removed from the dataset to avoid inaccuracy. Again, selective features were dropped from the data set since they were redundant in determining the instance. This led to reduction in the dimensions of the dataset resulting in 12 features and 1 output label out of the original 34 features and 1 label dataset. Furthermore, the dataset was highly skewed and imbalanced. To tackle this, the minority class were oversampled in order to maintain a balanced dataset for efficient model training and results. The authors were able to get the best accuracy (100%) using a multi-layer perceptron. A constructive and detailed analysis of all the literatures reviewed alongside their best performance score i.e., accuracy is prepared in Table.1

TABLE I. ANALYSIS OF LITERATURE REVIEW

Year	Dataset	Model	Accuracy
N. P. Surekha et al. [8] (2024)	Kaggle – "Asteroid Dataset"	Light Gradient Boosting with Importance	100%
M. Sam Kennanya et al. [9] (2023)	Jet Propulsion Laboratory	Random Forest with borderline SMOTE	99.86%
V. Bahel et al. [10] (2021)	Small-Body Database	Random Forest Classifier	99.99%
N. Ranaweera et al. [11] (2022)	Small Body Database (JPL HORIZON)	K–Nearest Neighbor (KNN)	91%
A. Sandeep et al. [12] (2024)	Self-procured	Decision Tree Classifier	99%
L. Almousa et al. [13] (2025)	Minor Planet Center (Kaggle)	Multi layered Perceptron	100%

III. Proposed Method

Different algorithms work differently and it is entirely up to us as to which algorithm we should train our model to obtain the best results. To get the desired result in this comparative and experimental study, supervised machine learning technique is applied which includes algorithms like Linear Regression, Logistic Regression, Ridge & Lasso Regression, Random Forest Classification, Naive Bayes Classifier, k-Nearest Neighbors. The are 8 subsections which describes the following: A. Dataset Used, B. Hardware and Software, C. Model Workflow, D. Data Visualization, E. Data Preprocessing, F. Feature Selection, G. Algorithm Used and H. Models employed with score.

A. Dataset Used

The dataset was taken from Kaggle [14] which was originally created and maintained by the Jet Propulsion Laboratory of California Institute of Technology, an organization under NASA. The dataset comprises of 958524 rows and 45 columns, all consisting of relevant asteroid information and useful and significant features in the field of astrophysics for predicting and classifying asteroids.

B. Hardware and software

Test simulations and experimental analysis were performed on a workstation running Windows 11 Home Single Language OS and with an Intel 12th Gen Intel(R) Core (TM) i7-1255U 1700 MHz CPU, 10 cores, 12 logical processors and 16 GB RAM. Scikit-Learn modules allowed for the usage of Python 3 version with Anaconda Spyder carried out in Jupyter Notebook.

C. *Model Workflow*

A well proposed model workflow must be designed in order to develop a robust machine learning model. The flowchart of all the necessary processes required to develop a model and predict and test it on a processed dataset can be seen in Figure-1.

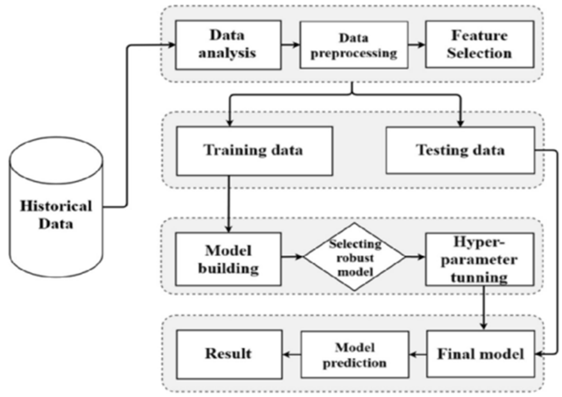

Fig. 1. Flowchart for the Prediction of Hazardous Asteroids

D. *Data Visualization*

The dataset can be visualized by the means of several data attributes such as the asteroid class which is described in Table-1 and Figure-2.

TABLE I. ASTEROID CLASS DISTRIBUTION

S. No	Class	Count	Orbit Class
1	MBA	855954	Main belt Asteroid
2	OMB	28355	Outer main belt
3	IMB	20360	Inner main belt
4	MCA	18685	Mars Crossing Asteroid
5	APO	12687	Apollo
6	AMO	8457	Amor
7	TJN	8221	Jupiter Trojan
8	ATE	1729	Aten
9	CEN	506	Centaur
10	AST	76	Other Asteroid
11	IEO	22	Interior Earth Object
12	HYA	4	Hyperbolic Asteroid

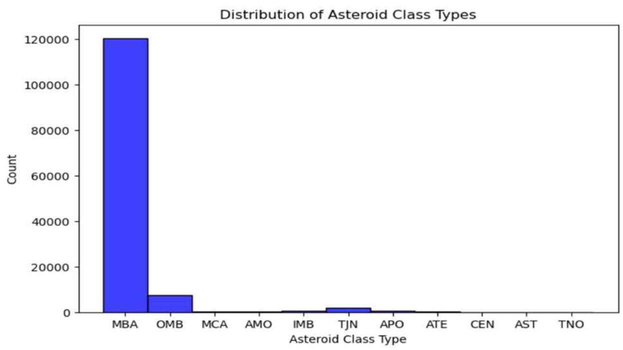

Fig. 2. Histogram of Asteroid Class Types

E. *Data Cleaning & Processing*

It is an essential technique which must be implemented prior to the development of the model as they must be trained and developed on a consistent dataset, devoid of redundant information which does not contribute in the fulfillment of the model's objective, which is to correctly predict the hazardous nature of asteroid in this case. The dataset also consisted of some null attributes which were truncated in order to avoid performance degradation of the training process. The data set which was used by the machine learning models consisted of now, 131142 rows and 32 columns. The heat map of correlation of all the relevant attributes post data processing has been plotted as shown in Figure-3.

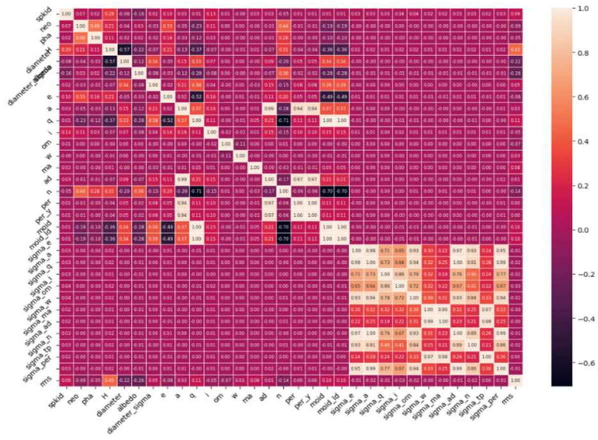

Fig. 3. Correlation Heatmap of all features after processing and cleaning the dataset

F. *Feature Selection*

In order to optimize our dataset, the most significant, important and relevant columns or features required to achieve our goal were selected and then modified. They were - PHA flag and NEO flag. Originally, they consisted of 2 values: Yes or No. It was understood and evident that the machine learning models work best on numeric data and even better on binary data so accordingly, the values of both the flags were updated and replaced as follows - Yes was replaced with 1, No was replaced with 0. This ensures a decrease in the margin of error of the models and assists greatly in its performance.

G. *Algorithms Used*

After the data was processed, cleaned and the process of feature selection was completed and wrapped up, it was time for the development of our model. Few of the supervised machine learning algorithms were implemented and further trained and tested to classify the asteroids as hazardous or non-hazardous. These algorithms can further be classified into two types - Regression algorithms and Classification algorithms. Regression algorithms that were used were namely - Linear, Logistic, Ridge & Lasso, along with classification algorithms such as Random Forest, Naive Bayes, k-Nearest neighbor. A crucial process prior to the creation and development of model was to make the training datasets and testing datasets. The dataset was divided into 2 data sets - x & y, where x was the dataset devoid of the output labels ('pha' and 'neo') and y was the dataset containing only the output labels. Further 'x' and 'y', individually, were classified into training and testing dataset. Necessary conversions were carried out in the classified datasets to ensure every value was of only numeric datatype. All the columns in the dataset along with their datatype has been plotted.

1) Linear Regression

It is a statistical model established to model the relationships between dependent variables and multiple independent variables. Using labeled data points, it maps the data points that are helpful for making predictions on a fresh dataset. It tries to compute a linear relationship between the dependent and non-dependent variable by 'fitting' a linear equation and provides a continuous output based on the input variables. This linear equation is widely known as best-fit line which demands that the error between the predicted data points and actual data points should be kept to a level minimum. In machine learning, the intercept of the best-fit line is known as bias and the slope is known as the weight of the line.

2) Lasso Regression

Lasso regression, shortened version of Least Absolute Shrinkage and Selection Operator, is a dynamic mix of supervised machine learning algorithm and a statistical regularization technique which is used for handling large datasets and predicting target values. It is widely used to prevent overfitting. It is a phenomenon that occurs if the model performs well with the training data (known as low bias) but fails to perform well with the test data (high variance).

3) Ridge Regression

Ridge regression (also known as R2 regularization) is a similar regularization technique as Lasso, the only difference being that it repeatedly attempts to shrink the coefficient to zero when the penalty term is raised to a higher value which in turn lowers the variance of the model.

4) Logistic Regression

Logistic regression aims to solve classification problems. It is a flexible and robust approach as it solves problems by predicting categorical outcomes, instead of continuous outcome. It is widely used in the process of determination the probability of predicting an instance is a part of a class or not. A sigmoid function maps the independent variable into a value between 0 and 1.

5) Random Forest Classifier

This method can be applied to both, regression and classification problems. It predicts categorical outcomes based on the input by utilizing multiple amounts of decision trees and outputs the label which obtained the maximum count of votes by all the individual trees. This algorithm was implemented as it is known for handling large and noisy data and the ability to yield optimal results, tackling the problem of class imbalance which was a major issue in the data set.

6) Naive Bayes Intuition

This algorithm works on the basis of probability which is useful for classification and regression purposes. It operates solely on the basis of Bayes' theorem. The importance of this theorem cannot be understated when it comes to probability prediction of certain values. Using the features from new data points, it can calculate the probability of each class. Ultimately, the class with the highest probability value is classified as the prediction of the new data.

7) K-Nearest Neighbor Classifier

Also abbreviated as KNN, this algorithm makes use of the neighboring data points to train the model and make predictions. k is a numeric value, which indicates the number of neighboring data points the model will take into consideration when taking necessary actions at the time of the classification and making decisions afterwards.

H. Models Employed with Score

All of the models mentioned and discussed above were practically implemented to predict the outcome which is to see and analyze whether an asteroid can be regarded as hazardous or not. Subsequently, each of the models were tested out and following are the results that were obtained.

a. Linear Regression

Classic regression metrics were evaluated after training the model and testing it to make predictions. The scores obtained can be seen in the following

TABLE II. METRIC SCORE OF LINEAR REGRESSION

RMSE Loss	**R^2 Score**
0.046377	0.287816

b. Ridge Regression

The above-mentioned metrics were also evaluated for the Ridge regression model which is as follows in Table.III.

TABLE III. METRIC SCORE OF RIDGE REGRESSION

RMSE Loss	**R^2 Score**
0.046385	0.287734

c. Lasso Regression

Again, the examination of the model was carried out on the basis of the evaluation metrics used in Table. IV:

TABLE IV. METRIC SCORE OF LASSO REGRESSION

RMSE Loss	**R^2 Score**
0.058057	0.287734

d. Logistic Regression

Since multiple columns are under training, implementing a multi-level classifier to our logistic regression model is just as mandatory. Following scores were obtained (Table.V.) after the evaluation of performance metrics.

TABLE V. METRIC SCORE OF LOGISTIC REGRESSION

Accuracy	**Precision**	**Recall**	**F1 Score**
0.998322	0.916167	0.831521	0.871794

e. Random Forest Classifier

The model was trained using a couple of parameters listed below and was then assessed on the basis of performance metrics as follows:

n_estimators - It indicates the total amount of trees that were used and trained for the model. It was set to 1000. Though it greatly increases the computation efficiency, it increases the computation time;

class_weight - This parameter was used mainly to adjust the weights of the classes. It was set to balance to minimize class imbalance.

The following evaluation metrics were calculated for the micro averaging parameter for Random Forest classification, shown in Table.VI.

TABLE VI. METRIC SCORE OF RANDOM FOREST

Accuracy	Precision	Recall	F1 Score
0.994565	0.994565	0.994565	0.994565

f. Naive Bayes Classifier Intuition

A Gaussian distributive approach was used in the development of the model. Such an approach was chosen on the basis of the type of dataset the model was trained with. The data consists of binary and real-valued attributes, and since many input labels were continuous and not discrete, a Gaussian (normal) distribution was preferred. Following are the results obtained, as shown in Table.VII:

TABLE VII. METRIC SCORE OF NAIVE BAYES

Accuracy	Precision	Recall	F1 Score
0.99294	0.53179	0.994565	0.685354

g. K-Nearest Neighbor Classification

Like any other model, performance metrics were measured and evaluated and plotted in a tabular format as shown in the Table VIII below. Value of k is taken as 1 in this case.

TABLE VIII. METRIC SCORE OF NAIVE BAYES

Accuracy	Precision	Recall	F1 Score
0.99295	0.808080	0.43478	0.563771

IV. Results And Discussion

This study aimed to accurately identify the type of an asteroid using its supervised learning algorithms while improving predictive accuracy on multiple statistical parameters. The dataset contained features of asteroid behavior, both numerical and categorical variables.

These were all of the models that were trained on several different algorithms and a diverse range of results and outcomes were obtained. A comparative study and analysis was done on all the algorithms the model was training, testingand making decisions on. The comparison of the results is constructed in a tabular format for a better in-depth understanding.

Also, for optimal comparison and result analysis, the best of all averaging parameter is considered for every model which has been trained upon. The table clearly analyses all the performance scores of every classification algorithm the model was trained upon. Each model was evaluated using four different classification metrics namely Accuracy, Precision, Recall and F1 score.

It is evident from the analysis that Random Forest Classifier proved to be the most suitable algorithm in predicting the true nature of asteroids. By achieving the highest accuracy (99.99%), it suggests that almost every prediction was made correct indicating excellent and robust performance across the dataset. Using this algorithm, the model handled both categorical and numerical features gracefully and combined multiple decision trees to reduce overfitting.

On the other hand, Logistic Regression also had a decent accuracy but low recall meaning that it missed some portion of true positives. Naïve Bayes has a low precision rate meaning that it predicts almost everything has positive which in turn increases the count of false positives. KNN struggled with overfitting and feature scaling thus leading to poor Recall and F1 score.

Following is the comparative analysis between the algorithms model was trained on and the performance metrics of each algorithm.

Further, following bar graphs were plotted as shown in Figure-4 using Matplotlib which were then grouped on the basis of the performance metrics of each mode. The visualization provided valuable insights for analyzing and comparing the best classification model. Table IX displays the performance scores of classification models in terms of Accuracy, Precision, Recall and F1 Score. Table X displays the evaluation scores of all three regression models in terms of RMSE and R2 Score.

TABLE IX. PERFORMANCE SCORES OF CLASSIFICATION MODELS

Model	Accuracy	Precision	Recall	F1 score
Logistic Regression (micro)	0.99832	0.91617	0.83152	0.8718
Random Forest (micro)	0.99992	0.99456	0.99456	0.9945
Naive Bayes (weighted)	0.99295	0.53179	0.99457	0.6853
KNN (k = 1, micro)	0.99295	0.80808	0.43478	0.5653

TABLE X. EVALUATION SCORES OF REGRESSION MODELS

Model	RMSE	R2 Score
Linear Regression	0.046377	0.287816
Ridge Regression	0.058057	0.287734
Lasso Regression	0.058057	0.287734

The Linear Regression model has the lowest RMSE score as compared with other models. The R2 score is almost similar of all regression models. The consolidated bar graph with the accompanying comparison table captures visually and numerically how each model performed across all four metrics. It can be concluded strongly that out of all the classification models trained, Random Forest Classifier triumphed over other models in all aspects of performance metrics and gave an overall balanced score for each of them.

However, it comes with a computational overhead and costly training time due to the high number of trees involved in the process. In some restricted cases like initial screening models or constrained environments, Logistic Regression or Naïve Bayes classifier can also be used since they are faster, interpretable and yield a high recall rate.

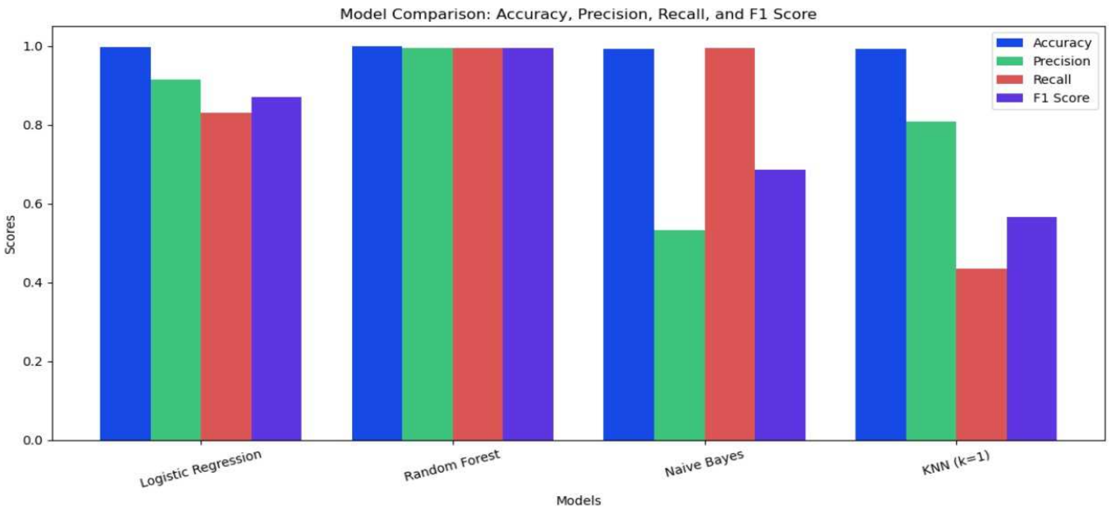

Fig. 4. Model Comparison

V. Conclusion & Future Scope

The field of astronomy and astrophysics is constantly expanding and becoming a crucial segment in the study and analysis of celestial bodies. This leads to the possibility of existence of unknown objects and particles which might pose as a risk to life in Earth. Dependable models are required more than ever before in determining the threats human life could possibly face in the near future.. Along with the perfect model and algorithm, the first and arguably, the most important part in implementing such a task is to gather, collect, organize and process the data which will be supplied to the model. The dataset from Kaggle was cleaned and processed leading to elimination of inconsistent and irrelevant features which could pose as a liability and decrease the model's performance and efficiency. The determination of whether an asteroid was hazardous or not was made possible in large part by the application of supervised machine learning algorithms. It was clear and evident that out of all the algorithms the model was trained on: Linear Regression, Ridge Regression and Lasso Regression, Multi-output Logistic Regression, Random Forest Classification, k-Nearest neighbor and Naïve Bayes. The most accurate and precise algorithm was Random Forest Classification. With an accuracy of 99.99%, it can be implied that this algorithm was the best suited for our purpose and goal, which was to accurately predict the hazardous nature of asteroid and determining whether it is threatening to our planet Earth or not. Further, this field and area of research significantly provides more opportunities for further exploration and analysis. Scientists and engineers can make use of image analysis which can greatly enhance the study of PHA which will lead to more accurate results and make room for more valuable insights in the field of astronomy. In future, we can also apply all the models on different datasets and can also go for deep learning approach.

References

[1] S. Khudikyan, A. Chamberlin, "NEO Basics", National Aeronautics and Space Administration, http://cneos.jpl.nasa.gov/about/nea_resource.html

[2] H. Shang, X. Wu, D. Qiao, X. Huang, Parameter estimation for optimal asteroid transfer trajectories using supervised machine learning, Aerospace Science and Technology, Volume 79, pp. 570-579, 2018

[3] T. Sharma, S. Sharma, A. Sharma, A. Kumar, A. Malik and A. Sharma, "Asteroid Hazard Prediction Using Machine Learning: A Comparative Analysis of Different Algorithms," 2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE), GHAZIABAD, India, pp. 673-677, 2023

[4] V. Carruba, S. Aljbaae, R. C. Domingos, A. Lucchini, P. Furlaneto, Machine learning classification of new asteroid families members, *Monthly Notices of the Royal Astronomical Society*, Volume 496, Issue 1, pp. 540–549, 2020.

[5] T. Choudhury, D. Ghosh, M.K. Gourisaria, J.J. Jena, P. Pattnayak, A. Bandyopadhyay. "Revolutionizing H5_HPAI Detection: Role of Machine Learning in Early Diagnosis." In 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), pp. 882-887. IEEE, 2024.

[6] S.K. Bharti, D. Ghosh, M. K. Gourisaria, J. J. Jena, P. Pattnayak, S.S. Patra. "Beyond the Biopsy: A Comprehensive Machine Learning Based Approach to Thyroid Cancer Staging." In 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), pp. 846-851. IEEE, 2024.

[7] Y. Singh, D. Ghosh, M.K. Gourisaria, J.J. Jena, S.S. Patra, A.R. Panda. "Machine Learning Empowered Osteoporosis Prediction: A Comparative Analysis." In 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 962-967. IEEE, 2024.

[8] N. P. Surekha, P. Nanjudan, S. Bashir, "Prediction of Hazardous Asteroids using Machine Learning" 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC), pp. 2, 2024

[9] M. Sam Kennanya, T. Meena, M. Sai Pravardhitha, A. Sai Vignesh, "Classification of Potentially Hazardous Asteroids Using Artificial Neural Networks and Over Sampling Techniques" 2023 Global Conference on Information Technologies and Communications (GCITC), pp. 6, 2023

[10] V. Bahel, P. Bhongade, J. Sharma, S. Shukla, M. Gaikwad, "Supervised Classification for Analysis and Detection of Potentially Hazardous Asteroid" 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), pp. 1-4. Dec. 2021

[11] R. N. Ranaweera, TGI Fernando, "Prediction of Potentially Hazardous Asteroids using Deep Learning" 2022 2nd International Conference on Advanced Research in Computing (ICARC), pp. 1-6. Feb. 2022

[12] A. Sandeep, P. J. Reddy, T. M. Perkin, "PLANETARY DEFENSE USING MACHINE LEARNING EARLY WARNING SYSTEMS FOR HAZARDOUS ASTEROIDS" 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), pp. 1-6. Apr. 2024

[13] L. Almousa, L. Hamdan, T. Hammouri and H. Abdel-Nabi, "Asteroid Orbit Classification with Machine Learning: A Data-Driven Approach," 2025 International Conference on New Trends in Computing Sciences (ICTCS), Amman, Jordan, 2025, pp. 156-163

[14] M. S. Hussain, "Asteroid Dataset" Kaggle 2023, Retrieved on 16th Jan 2025 https://www.kaggle.com/datasets/sakhawat18/asteroid-dataset/data

[15] D. Khajuria, A. Sharma, N. Sharma, M. Mangla, "Classification and Comparative Analysis of Earth's Nearest Objects using Machine Learning Models" 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1-8. Mar. 2023