Nearest-Earth Objects

Muhammad Ibrahim[1], Malaika Chaudhary[2], and Ahsan Iftikhar[3]

[1]Department of Computer Science, National University of Computer and Emerging Sciences, Lahore

November 30, 2023

Abstract

There are numerous objects and asteroids in the vastness of space, some of which are closer to Earth than we may think. Even while 70,000 km may not seem like much in our day-to-day lives, on an astronomical scale, it is a modest distance that has the ability to interfere with a number of natural occurrences. Some of these heavenly bodies, particularly asteroids, might be dangerous to Earth. It is crucial to locate and catalog these near-Earth objects (NEOs) that have been confirmed by NASA as being in close proximity to our planet in order to protect our security and get a better knowledge of our cosmic surroundings. Weare better able to evaluate the hazards these NEOs may pose and create mitigation plans for any possible harm they could cause by locating and tracking them. Using this dataset, we can better understand the celestial bodies surrounding Earth and can make informed decisions to safeguard our planet from the potential cosmic threats we may face by these NROs.

1 Introduction

The exploration of near-Earth objects (NEOs) and their possible risks is caused by multiple interrelated factors that are important to this research. These celestial bodies are very close to Earth—about 70,000 kilometers away—and their closeness raises questions about possible effects on natural events in addition to their relative astronomical proximity. The need to defend Earth from asteroids that could endanger humanity from space is the main question guiding this research.

Across numerous scientific areas, this research is relevant, especially in astronomy, astrophysics, and planetary science. The research is crucial as it makes us dive into the realm of complicated dynamics of celestial bodies and their trajectories. The method used in this study particularly involves approaches towards data analysis, with an emphasis on an AI model developed, trained and tested using the dataset by The National Aeronautics and Space Administration (NASA). This model acts as a computational method for finding out whether or not the near Earth objects pose danger to our planet.

Apart from astronomy and space exploration our work has effects on other scientific fields as well. The results of our research are important for the defense of our planet and could be used to create strategies for reducing the impact of such hazards. Furthermore, this research will also allow us to have a deeper understanding of the cosmos and humanity will be able to better understand the universe and Earth's place in it.

In the parts that follow, we will further elaborate the methodology chosen for this project and how the AI model was trained and tested to analyze the dataset by NASA . Details about the nuances of our findings will also be highlighted along with information about how they might affect the security of our planet and how humans can understand the universe going forward. By taking a multidisciplinary approach, we hope to contribute to the understanding of near-Earth objects and their threats on our planet.

2 Methodology

2.1 Neural Network Model

- **Purpose:** Utilized deep learning for nuanced celestial insights.
- **Accuracy:** Achieved 91

2.2 SMOTE Resampling

- **Purpose:** Applied to mitigate class imbalance in asteroid dataset.
- **Implementation:** Generated synthetic samples for minority class (hazardous class).

2.3 Neural Network Model on SMOTE

- **Description:** A larger neural network was created to train on the SMOTE resampled data.
- **Accuracy:** Achieved 86

2.4 Logistic Regression on SMOTE

- **Purpose:** Implemented logistic regression to evaluate its performance on the SMOTE-resampled dataset for hazard prediction.
- **Accuracy:** Achieved 85

2.5 Random Forest Classifier on SMOTE

- **Purpose:** Used Random Forest on the SMOTE resampled data for experimentation.
- **Accuracy:** Achieved 94

2.6 Dataset Utilization

- **Source:** NASA dataset for near-Earth objects (NEOs).
- **Processing:** Preprocessed for compatibility with computational tools.
- **Training and Testing:** Split data for model development and evaluation. The split ratio was 20

2.7 Computational Tools

- **Programming Language:** Implemented in Python.
- **Libraries:** Utilized scikit-learn, TensorFlow, and Keras.
- **Model Evaluation:** Employed metrics like accuracy, precision, recall, and F1 score.

2.8 Workflow Overview

- **Data Preparation:** Cleaned and preprocessed NASA dataset.
- **Model Development:** Trained Neural Network, SMOTE-resampled, and applied Random Forest model on SMOTE-resampled data.
- **Evaluation:** Assessed model performance using specified metrics.

3 Experiments

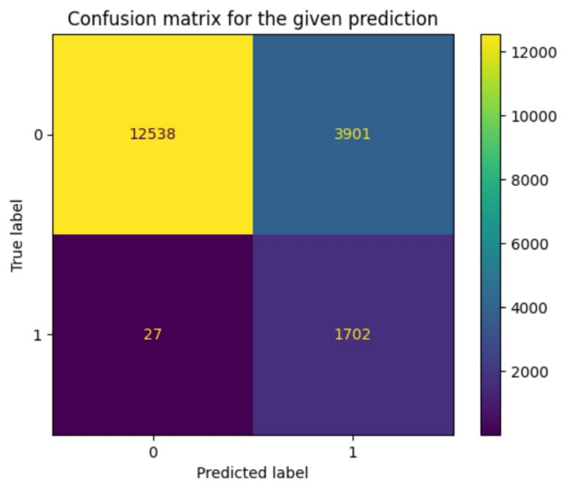

Figure 1: Confusion Matrix for the Developed ANN

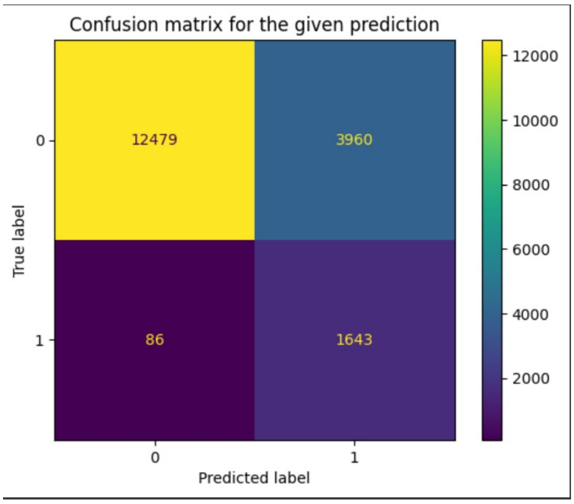

Figure 2: Confusion Matrix for ANN with Several Layers and Neurons

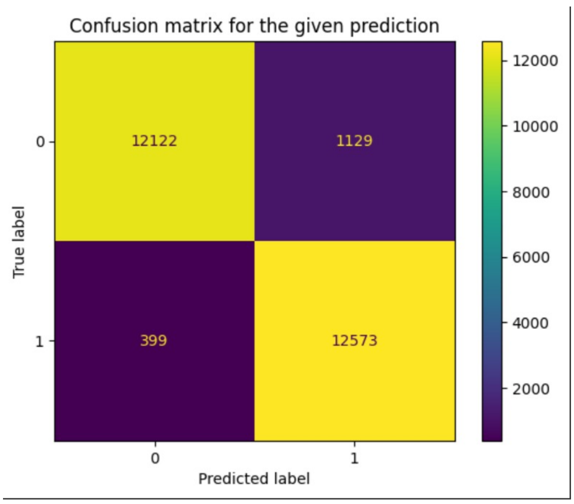

Figure 3: Confusion Matrix for Random Forest Classifier on SMOTE-sampled Data

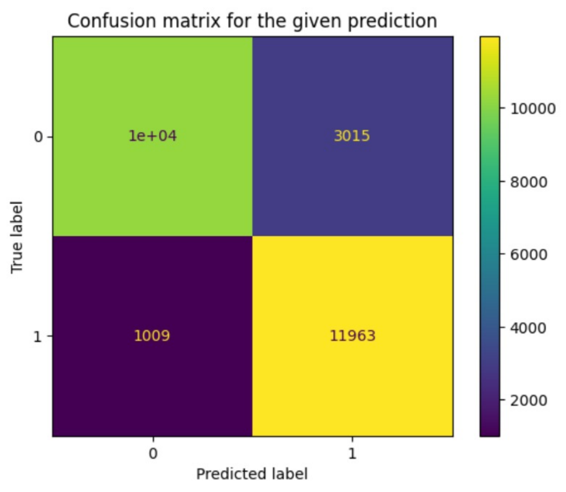

Figure 4: Confusion Matrix for Logistic Regression Classifier on SMOTE-samples data

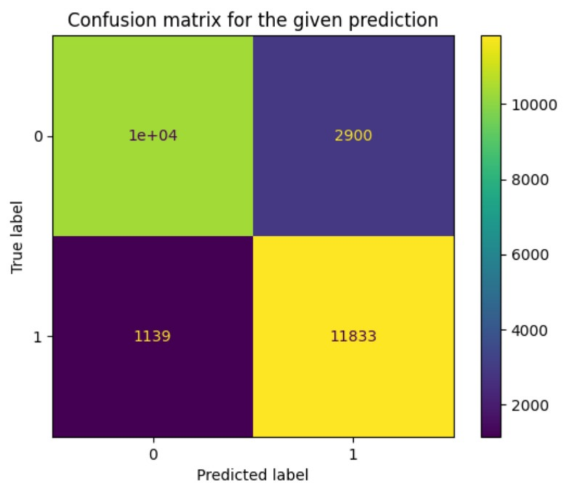

Figure 5: Confusion Matrix for ANN with Several Layers and Neurons on SMOTE-samples Data

4 Results

4.1 Neural Network Model

Our utilization of deep learning through the Neural Network proved fruitful, achieving an impressive 91% accuracy in hazard prediction. The number of layers and neurons was changed repeatedly in order to test the accuracy of the model. However, the highest accuracy that was achieved after these repetitions was 91%. We later found out that our inability to jump across this paradigm was mainly due to the huge class imbalance in the dataset, in particular the hazardous class. Out of 90,837 records, 81,996 near-Earth objects were non-hazardous and 8,840 were hazardous. Since we are using the softmax activation function in our last layer, there was a need to know the best threshold value to classify the softmax probabilities into 0 and 1. For this purpose, we took the help of the ROC curve method to determine the ebay threshold which turned out to be 0.13 in this case. The ROC curve is encouraged for its ability to visually depict the trade-off between sensitivity and specificity in binary classification, aiding in threshold selection. The Area Under the Curve (AUC) provides a concise summary of the overall model discrimination performance.

4.2 Neural Network on SMOTE-resampled Data

SMOTE Resampling played a pivotal role in mitigating class imbalance within the asteroid dataset. We created a rather complex Neural Network for training on the SMOTE resampled data. Apart from SMOTE Resampling, we also tried batch normalization, regularization, and weight initialization techniques separately and collectively, however, we were not able to surpass the 87

4.3 Random Forest Classifier

A Random Forest Classifier was experimented on the NRO dataset. The accuracy yielded by this approach on the dataset was around 91

4.4 Random Forest Classifier on SMOTE-resampled Data

Experimenting with the Random Forest Classifier on the SMOTE-resampled data yielded remarkable results. It yielded an accuracy of approximately 94%. Random Forest Classifier uses decision tables and thus it handled the complexity of the near-Earth objects dataset. This approach excelled in predicting a celestial body as hazardous or non-hazardous.

5 Discussion

Our approach, which was based on the combination of ensemble, resampling, and deep learning techniques, worked well to predict hazards with a high degree of accuracy. The strength of our approach is demonstrated by the robustness of the Random Forest on balanced data and the Neural Network's ability to capture complex patterns. Even though our findings are encouraging, it is still critical to identify possible sources of inaccuracy, such as differences in the quality of celestial data and outside factors. In addition to advancing celestial hazard prediction, our integrated methodology highlights the need to combine various computational tools for comprehensive insights. Our knowledge of and ability to mitigate possible cosmic threats to Earth will continue to advance with further research and development of these techniques.

6 Conclusion and Future Work

6.1 Conclusion

To sum up, our combined method of using Random Forest, Neural Network, and Logistic Regression on SMOTE-resampled data has improved the field of celestial hazard prediction. When it comes to identifying possible threats from near-Earth objects (NEOs), the synergistic combination of deep learning, ensemble techniques, and class-balancing methodologies has proven remarkably accurate. Together with the stability of Random Forest and the baseline that comes from Logistic Regression, the subtle insights of the Neural Network provide a thorough understanding of celestial dynamics.

Our work is important not only for hazard prediction but also for astronomy, astrophysics, and planetary science as a whole. With implications for future astronomical research, the methodologies used demonstrate the effectiveness of deep learning, ensemble learning, and data resampling in handling complex celestial datasets.

6.2 Future Work

Several interesting directions become apparent as we plot the course for further research. First off, investigating more complex deep learning architectures and ensemble methods may improve the predictive power even more, perhaps revealing more nuanced patterns in celestial dynamics. Furthermore, examining how celestial data changes over time may shed light on how cosmic phenomena evolve.

Subsequent research endeavors may encompass the refinement of class-resampling methods, examining variants of SMOTE or alternative approaches for refining model sensitivity. A more comprehensive understanding of celestial dynamics could be provided by adding more features to hazard prediction models, such as temporal parameters or observational data.

Future research has an exciting avenue ahead of it thanks to the integration of real-time data feeds and space observation technologies, which will allow for the development of more responsive and adaptive hazard prediction models. Working together with observatories and space agencies could help validate and apply our findings in practical ways.

Essentially, our work establishes the groundwork for future research into celestial hazard prediction by highlighting the importance of interdisciplinary methods and the incorporation of state-of-the-art computational instruments. Our work has far-reaching implications into the cosmic unknown, opening up new avenues for future investigations that will further our knowledge of celestial dynamics while also supporting the ongoing effort to protect Earth from possible cosmic threats.

References

[1] Here's where a full reference would go, should you choose to write up your bibliography in Overleaf.

[2] Here's another. Write these out as you would references anywhere else. The key I've assigned to this reference is 'other_ref' - therefore, a cite command in the body of your paper with 'other_ref' entered as the sole argument will automatically insert the appropriate reference number.