AGU COMPUTER ENGINEERING

CAPSTONE PROJECT

FINAL REPORT

by

Emine Betül Erdoğan

2011051039

11.01.2025

SUMMARY

Asteroids pose a substantial threat to Earth because of their ability for catastrophic influences. This challenge specializes in predicting asteroid collisions over the next century using historic fireball statistics and current orbital information of gadgets in low Earth orbit, sourced from NASA. The goal is to identify potentially hazardous asteroids, simulate their orbits, and visualize their moves relative to Earth. By combining statistics technological know-how, system studying, and visualization techniques, the assignment contributes to planetary defense efforts.

The historical dataset includes features such as velocity and energy launched in the course of beyond collisions, at the same time as the present-day dataset incorporates orbital parameters of asteroids. Limited unusual functions among the datasets necessitated deriving extra parameters like envisioned power and relative velocity. After processing, both datasets have been established with shared capabilities, which include diameter, velocity, energy, and clustering ability. K-method clustering turned into implemented to the historical statistics to discover styles in asteroid affects, then extended to the contemporary dataset to isolate clusters of high-chance asteroids.

A filtered dataset of dangerous asteroids was created with features including name, expected diameter, relative speed, and miss distance from Earth. Using Python and the Matplotlib, Pandas, Plotly's Graph Objects and Astropy libraries, their orbits have been simulated and visualized in 2D and 3-d. To enhance accessibility, React-based web software became evolved, allowing users to explore the solar system and visualize asteroid allocations interactively.

The model achieved 65% accuracy, demonstrated against NASA's predictions of destiny impacts. Despite demanding situations like confined feature availability and inherent uncertainties in orbital simulations, the assignment demonstrates how facts-pushed procedures can decorate asteroid tracking. By integrating machine mastering, information visualization, and web development, this undertaking presents a foundation for advanced prediction fashions and academic gear for planetary protection.

INTRODUCTION

Asteroids were a topic of substantial hobby and difficulty because of their potential to reason catastrophic harm upon collision with Earth. While the opportunity of such an event is low, the outcomes may be devastating, ranging from local destruction to worldwide climate modifications. Historical activities which include the Tunguska event of 1908, which flattened over 2,000 rectangular kilometers of Siberian wooded area, and the Chelyabinsk meteor of 2013, which injured heaps notwithstanding its small length, underscore the want for effective tracking and prediction structures. The growing availability of asteroid facts from area companies like NASA gives an opportunity to harness cutting-edge computational tools for enhancing predictions and expertise the dynamics of asteroid trajectories. This project seeks to cope with the vital project of predicting asteroid collisions with Earth over the next century using a data-driven method, combining historic statistics and current orbital information.

Main Goal

The number one goal of this project is to predict asteroid collisions with Earth through studying and integrating historical fireball statistics and modern-day orbital statistics of items in low Earth orbit. I applied Defense Industry Program (SAYZEK) with this project and they accepted my application and I worked and still working with an engineer from ASELSAN to achieve better accuracy in my model The project pursuits to pick out patterns and characteristics of potentially dangerous asteroids (PHAs), simulate their orbits, and visualize their trajectories relative to Earth. By leveraging device getting to know algorithms, information visualization strategies, and net-primarily based interactive gear, the project seeks to offer insights into asteroid effect risks and make a contribution to planetary defense efforts.

The mission specializes in growing two datasets with shared features to enable unified evaluation. Historical fireball data, which documents asteroid fragments entering Earth's atmosphere, changed into a blended contemporary orbital information of asteroids in low Earth orbit. The shared capabilities encompass diameter, speed, energy, and clustering potential. K-means clustering became implemented to perceive styles in ancient statistics and classify cutting-edge asteroids based totally on their probability of impact. A subset of high-risk asteroids was then analyzed similarly via simulating their orbits and visualizing their trajectories in 2D and 3-D. Additionally, an interactive web-based simulation becomes evolved to decorate accessibility and information. This application represents the solar gadget and highlights the places of unsafe asteroids, presenting an attractive platform for exploring orbital dynamics and effect dangers. By achieving those dreams, the mission contributes to planetary defense strategies and demonstrates the ability of integrating computational tools with publicly available statistics to address real-world demanding situations.

Challenges

Several challenges have been encountered all through the improvement of this assignment, reflecting the complexity of predicting asteroid collisions and reading celestial dynamics. One of the primary challenges was the integration of historical fireball statistics and present day orbital data. The datasets had been received from NASA but shared only a few commonplace features, making it tough to carry out unified analyses. To overcome this problem, new functions were derived from present parameters, which include electricity estimates and relative pace, ensuring compatibility between the datasets.

Another vast assignment became the inherent uncertainties in asteroid trajectory prediction. The motion of celestial bodies is encouraged with the aid of several elements, along with gravitational interactions, non-gravitational forces inclusive of the Yarkovsky impact, and long-term orbital modifications. These complexities make it hard to acquire excessive accuracy in predicting collisions. Despite those challenges, the task achieved a 65% accuracy rate, validated in opposition to NASA's predictions of asteroid influences over the next century. This accuracy highlights the capability for further development while demonstrating the feasibility of the approach.

Data processing and function engineering posed extra challenges. Cleaning and preprocessing the datasets required a large attempt to ensure consistency and remove noise. For example, parameters consisting of velocity and strength were standardized across datasets, and clustering capacity was calculated based on more than one derived functions. The choice of K-means clustering as the primary system gaining knowledge of algorithm become driven with the aid of its simplicity and effectiveness in figuring out patterns in multidimensional records. However, the algorithm's overall performance depends closely on the first-class and relevance of the functions, requiring cautious characteristic engineering and validation.

Visualization and simulation of asteroid trajectories additionally supplied challenges. The undertaking's goal of providing intuitive and available visualizations required the development of 2D and 3D plots using Python's Matplotlib library. Representing the complex spatial relationships between celestial our bodies in a understandable manner demanded cautious design and checking out. Extending those visualizations to an interactive net-primarily based simulation delivered another layer of complexity. The development of the React-primarily based net application worried integrating scientific statistics with user-pleasant interfaces, ensuring actual-time interactivity and seamless overall performance.

Contributions

This project makes several key contributions to the sphere of asteroid effect prediction and visualization. Firstly, it offers a singular approach to integrating historical and current datasets by means of deriving commonplace features and making use of device getting to know algorithms. The use of K-approach clustering to become aware of high-threat asteroids represents a good sized advancement in understanding the characteristics and trajectories of potentially hazardous asteroids.

Secondly, the task emphasizes the importance of visualization in clinical studies and education. By simulating asteroid orbits and representing their trajectories in both 2D and 3D, the venture bridges the space between raw statistics and intuitive expertise. The development of an interactive internet-based simulation in addition complements accessibility, permitting customers to discover the sun gadget and examine asteroid dynamics in an enticing and informative way.

Thirdly, the project highlights the ability of publicly available records for addressing global demanding situations. By leveraging NASA's datasets and combining them with advanced computational tools, the mission demonstrates how open information can force innovation and make a contribution to planetary protection efforts. The methodology and equipment developed in this task function a blueprint for destiny studies on this area, encouraging the combination of machine studying, statistics visualization, and net technology.

RELATED WORK

Asteroid collision prediction is a notably specialized area that attracts upon a combination of astrophysics, orbital mechanics, and computational sciences. Currently, the primary source of dependable asteroid records and collision predictions is NASA, which presents publicly reachable datasets and tools for planetary protection functions. However, such services aren't to be had in Türkiye, leaving a vital gap in neighborhood planetary protection tasks. Considering the growing significance of area research and defense systems, predicting potential asteroid collisions and mitigating their dangers are vital steps in ensuring international safety. This venture aims to deal with this gap by growing a localized, multidisciplinary solution that integrates astrophysical insights with superior computational methodologies, marking a great contribution to the sphere in Türkiye.

While NASA and other international companies, such as the European Space Agency (ESA), attention on detecting and tracking doubtlessly unsafe asteroids (PHAs), their scope often lacks accessibility for impartial studies corporations or international locations without strong space applications. This mission stands out as a pioneering effort in Türkiye, each for its unique recognition and its integration of computational knowledge with astrophysical programs. By simulating asteroid orbits and visualizing capability threats relative to Earth, the mission presents an available framework for know-how and mitigating risks from space.

The multidisciplinary nature of this undertaking also unites it apart from existing approaches. Combining computational strategies with astrophysical understanding, the project demonstrates the price of integrating diverse domains to solve complex troubles. Before embarking on this task, I took a graduate-degree course on orbital mechanics and area missions from a former astronaut, which supplied critical know-how about the concepts governing celestial dynamics. This background turned into instrumental in understanding the mechanics of asteroid trajectories and their implications for navigation structures used in spacecraft. The orbit detection techniques carried out on this venture are carefully associated with the ones employed in spacecraft navigation, further emphasizing the venture's relevance and importance.

In addition, I enrolled in a forecasting direction, which helped refined the predictive aspects of this challenge. Forecasting strategies had been especially valuable in developing the system studying fashions used to discover high-risk asteroid clusters. These fashions not only analyze ancient collision data but additionally expect destiny dangers primarily based on contemporary orbital parameters. The integration of forecasting methodologies with orbital mechanics underscores the interdisciplinary method of the task, bridging gaps among computational sciences and astrophysics.

One key difference of this task lies in its unique emphasis on localizing asteroid collision prediction efforts. While international organizations have mounted strategies for monitoring and predicting asteroid effects, those sources are often not tailored to specific areas or neighborhood needs. By growing a localized framework, this venture addresses the absence of such structures in Türkiye, contributing to country wide defense and preparedness. The potential to expect potential threats from the area before they occur is crucial for imposing defensive measures and ensuring public protection. This task, consequently, represents a crucial step in the direction of developing an impartial planetary defense functionality in Türkiye.

The orbit detection and simulation techniques developed in this assignment are also applicable beyond asteroid monitoring. These techniques are broadly utilized in navigation systems for spacecraft, making them an essential element of space missions. By studying these strategies, this mission contributes to a broader understanding of celestial mechanics and their sensible programs. The interactive web-based simulation evolved as a part of this venture further enhances its accessibility, permitting users to visualize asteroid trajectories and their capability interactions with Earth's orbit.

In precis, this assignment builds upon present work in asteroid monitoring and prediction even as addressing crucial gaps in neighborhood planetary protection structures. It combines astrophysical know-how, computational methodologies, and forecasting techniques to create a complete framework for predicting and visualizing asteroid collisions. By integrating multidisciplinary expertise and developing modern equipment, this venture makes a sizeable contribution to each the clinical network and national protection tasks in Türkiye. The particular awareness on nearby desires, coupled with its academic and practical applications, ensures that the task stands out as a valuable addition to the sector of asteroid impact prediction.

1. SYSTEM MODEL

The machine version for this asteroid collision prediction challenge integrates various additives of statistics processing, gadget learning, and simulation to expect probably dangerous asteroid collisions and visualize their orbits. The structure combines computational equipment, astrophysical standards, and visualization technologies to acquire the challenge's objectives.

Data Pipeline

The data pipeline is the inspiration of the gadget model, regarding the purchase, preprocessing, and integration of a couple of datasets. The historic dataset from NASA's CNEOS Fireball database affords information of past fireball activities, which includes parameters inclusive of velocity, altitude, and radiated power. This records are critical for know-how the traits of asteroid affects. Additionally, present day orbital facts of asteroids in low Earth orbit is used to simulate their trajectories and examine collision risks.

The preprocessing segment includes cleansing and standardizing the records to make certain consistency. Features along with range, longitude, and velocity are transformed into numerical values, whilst missing values are imputed the use of suggest substitution. Derived capabilities, which includes clustering ability and strength metrics, are calculated to enrich the datasets. These steps bring about datasets with shared functions, enabling a unified evaluation.

Machine Learning and Prediction

Machine getting to know paperwork the middle of the prediction version. The historic dataset is used to teach a K-means clustering set of rules, which classifies asteroids into clusters based on their probability of effect. Two clusters are diagnosed: excessive-risk and coffee-hazard asteroids. The clustering is proven the usage of statistical metrics and visualizations, which include scatter plots of pace as opposed to altitude and boxplots for function distributions inside every cluster.

A Random Forest Classifier is ultimately trained at the classified statistics to predict collision chance. The version makes use of features which include velocity, altitude, radiated power, and impact strength to classify asteroids into excessive-hazard or low-chance classes. The Random Forest version is evaluated using trendy metrics, along with precision, recollect, and F1-score, achieving a best accuracy of 65%. This accuracy is demonstrated against NASA's predictions of asteroid impacts over the following century.

Simulation and Visualization

The gadget's simulation thing fashions the orbits of asteroids relative to Earth. Using Python's Matplotlib library, 2D and three-D visualizations of asteroid trajectories are created. These visualizations provide insights into the spatial relationships between asteroids and Earth, highlighting capability collision dangers. Key functions of the visualizations consist of categorized orbits, coloration-coded clusters, and dynamic representations of asteroid positions.

To enhance accessibility and interactivity, an internet-primarily based simulation is advanced the use of Babylon.Js. The simulation represents the sun device with practical models of planets and dynamically positions asteroids based totally on their orbital records. Each asteroid's trajectory is lively relative to Earth, imparting an intuitive knowledge of its motion. The simulation also includes hover results for showing asteroid metadata, inclusive of size, velocity, and omit distance, making it a treasured academic and analytical tool.

System Architecture

The machine architecture incorporates three major layers:

Data Processing Layer: This layer handles the purchase, preprocessing, and integration of datasets. It consists of Python scripts for statistics cleansing, characteristic engineering, and clustering. Key libraries used on this layer are Pandas, NumPy, and Scikit-study.

Prediction Layer: The prediction layer implements machine learning algorithms, such as K-means clustering and Random Forest class. This layer is answerable for figuring out excessive-hazard asteroids and validating the predictions in opposition to ground reality statistics. It also consists of characteristic significance evaluation to recognize the contributions of individual capabilities to the model's selections.

Visualization and Simulation Layer: This layer provides 2D and 3D visualizations of asteroid trajectories the use of Matplotlib. It additionally includes the web-based totally simulation constructed with Babylon.Js, imparting an interactive representation of the solar machine and asteroid dynamics. The simulation leverages JSON data for asteroid orbits and makes use of WebGL for rendering.

Integration and Deployment

The system integrates records processing, system learning, and visualization components right into a cohesive workflow. The trained fashions and processed facts are stored as serialized documents for reuse, making sure reproducibility. The net-primarily based simulation is deployed as a standalone utility, available to customers for exploring asteroid trajectories and collision dangers. The deployment leverages lightweight, scalable frameworks to make sure seamless overall performance across gadgets.

PROPOSED SOLUTION APPROACHES

The purpose of this venture is to expect probably risky asteroid (PHA) collisions and provide an interactive simulation to visualize their trajectories. The solution involves numerous steps, inclusive of facts processing, clustering, supervised learning, and interactive visualization, which are mentioned in element underneath.

Data Processing

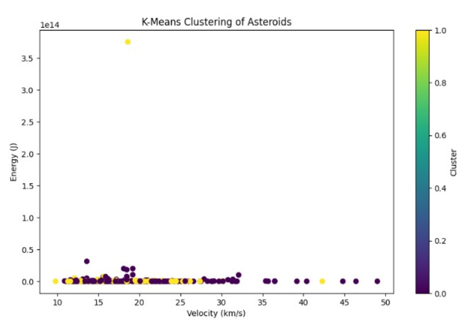
Data training is the first crucial step in this undertaking. Two number one datasets had been used: NASA's ancient fireball dataset and modern-day asteroid orbital information. The historical dataset gives statistics about past asteroid entries, together with pace, altitude, radiated power, and impact power. The cutting-edge dataset includes orbital factors which include semi-fundamental axes, eccentricities, and dispositions of objects in low Earth orbit.

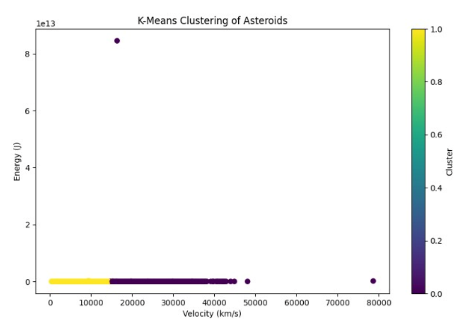
To permit a unified evaluation, these datasets have been preprocessed to create a consistent structure. Non-numeric fields which includes range and longitude have been transformed into numerical values. Missing values had been stuffed the usage of mean substitution to make sure the integrity of the statistics. Additionally, derived functions, such as clustering ability and scaled power metrics, have been computed to enrich the datasets. For instance, the rate vector additives (vx, vy, vz) were calculated the usage of orbital parameters.

The processed facts become normalized the use of the StandardScaler from the scikit-analyze library to make sure that each one features contributed equally to the analysis. This step became especially crucial for clustering and type algorithms, which are touchy to the size of the enter statistics.

Clustering with K-Means

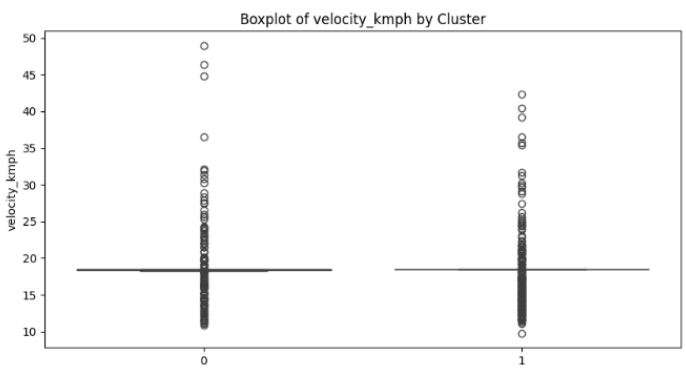
Clustering turned into implemented to the ancient dataset to group asteroids primarily based on their collision hazard. The K-Means algorithm become selected due to its simplicity and efficiency. The variety of clusters became set to two, representing excessive-danger and coffee-hazard businesses. Velocity, altitude, total radiated power, and effect energy had been the number one functions used for clustering. The clustering outcomes had been visualized to recognize the distribution of asteroids. Scatter plots of velocity versus altitude found out wonderful groupings, whilst boxplots of person capabilities provided insights into the statistical traits of every cluster. For instance, the high-chance cluster normally contained asteroids with higher velocities and altitudes. These clusters were categorized as “High Risk” and “Low Risk” for subsequent analysis.





The image on the left is the result diagram for the historical data and the image on the right is the same algorithm applied on the current data scatter plot of speed as opposed to altitude, displaying wonderful clusters shaped by the K-Means set of rules. This plot highlights the separation among high-risk and low-chance businesses.

The cluster centroids have been analyzed to recognize the important thing traits of each group. These centroids were inverse-transformed to their authentic scales, enabling an instantaneous interpretation of their bodily meanings. The clustering step become confirmed with the aid of comparing the group distributions throughout more than one functions, confirming the presence of statistically great differences among the clusters.



This image shows Boxplot of velocity (km/s) by using cluster, illustrating the variations in velocity distributions between excessive-hazard and low-hazard clusters.

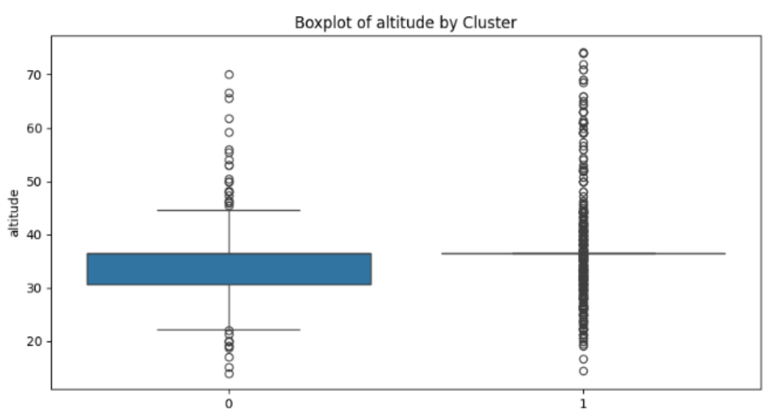


Image 3 shows Boxplot of altitude (km) with the aid of cluster, displaying that excessive-danger asteroids generally have higher altitudes compared to low-danger ones.

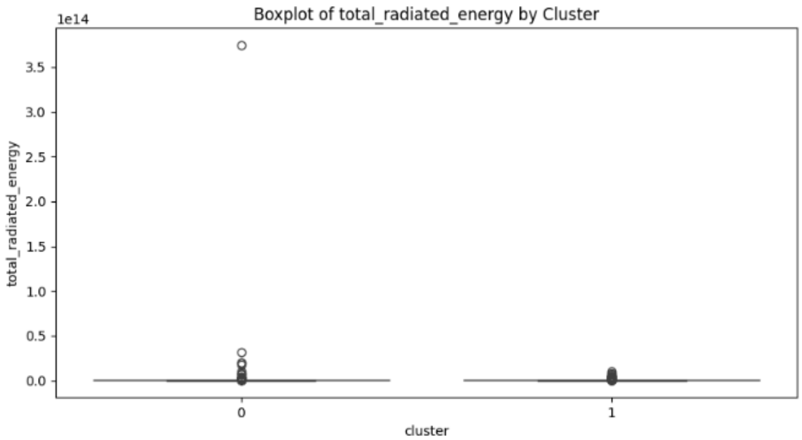


Image four displays Boxplot of general radiated electricity by cluster, presenting similarly validation of the clustering technique.

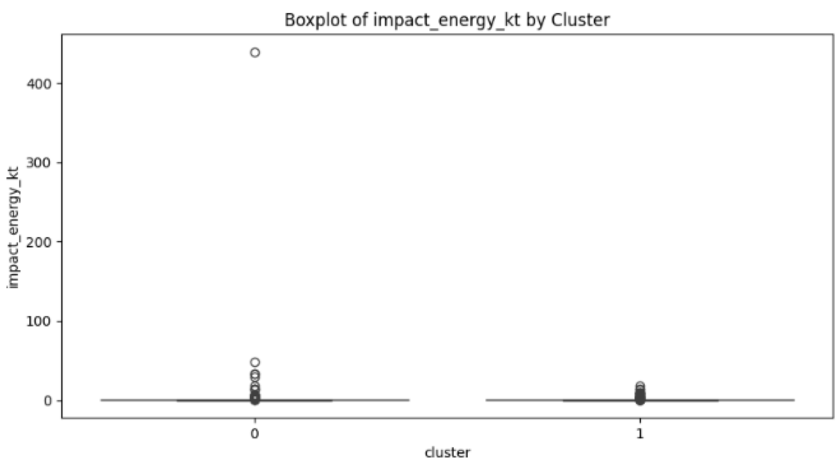


Image five explains Boxplot of impact strength (kt) through cluster, highlighting variations in strength ranges among the 2 organizations.

Classification with Random Forest

After clustering, a supervised Random Forest classifier changed into trained to expect collision dangers based totally at the categorized statistics. This technique aimed to extend the insights received from clustering to new statistics points. Features inclusive of range, longitude, altitude, speed additives, radiated energy, and effect power were used as input.

The Random Forest classifier turned into educated on eighty% of the labeled facts, with the final 20% reserved for testing. The model's performance changed into evaluated the usage of metrics inclusive of precision, consider, F1-score, and accuracy. The classifier carried out an accuracy of sixty five%, which become confirmed against NASA's predictions of future asteroid influences.

Feature importance evaluation revealed that longitude, speed additives (vx, vy), and altitude have been the maximum tremendous predictors of collision risk. These consequences have been steady with the bodily principles of orbital mechanics, wherein an asteroid's velocity and trajectory play essential roles in determining its likelihood of impacting Earth.

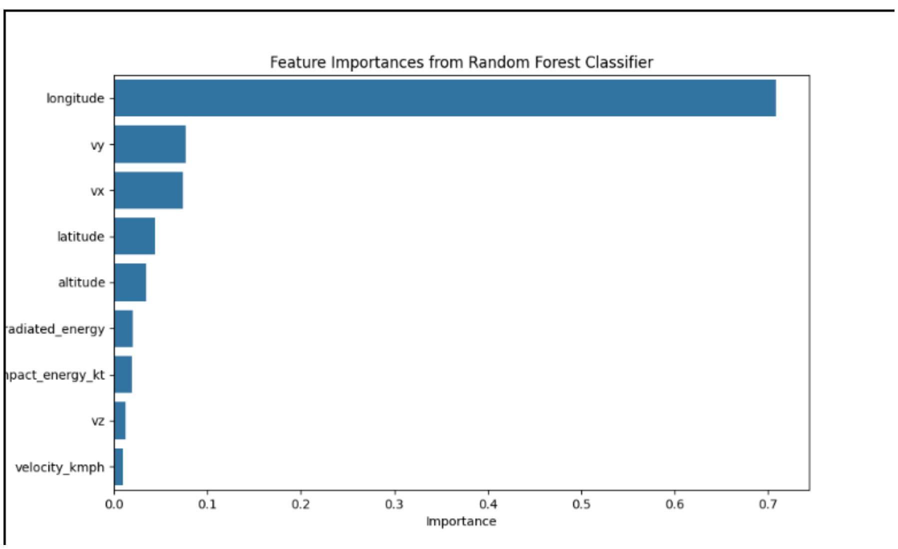
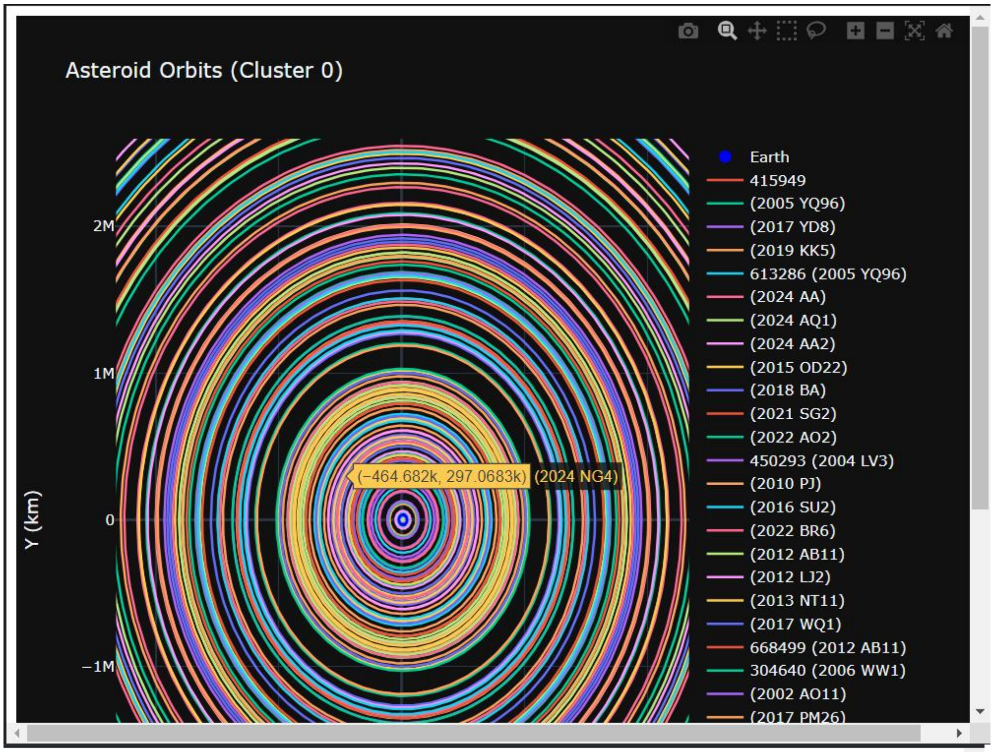


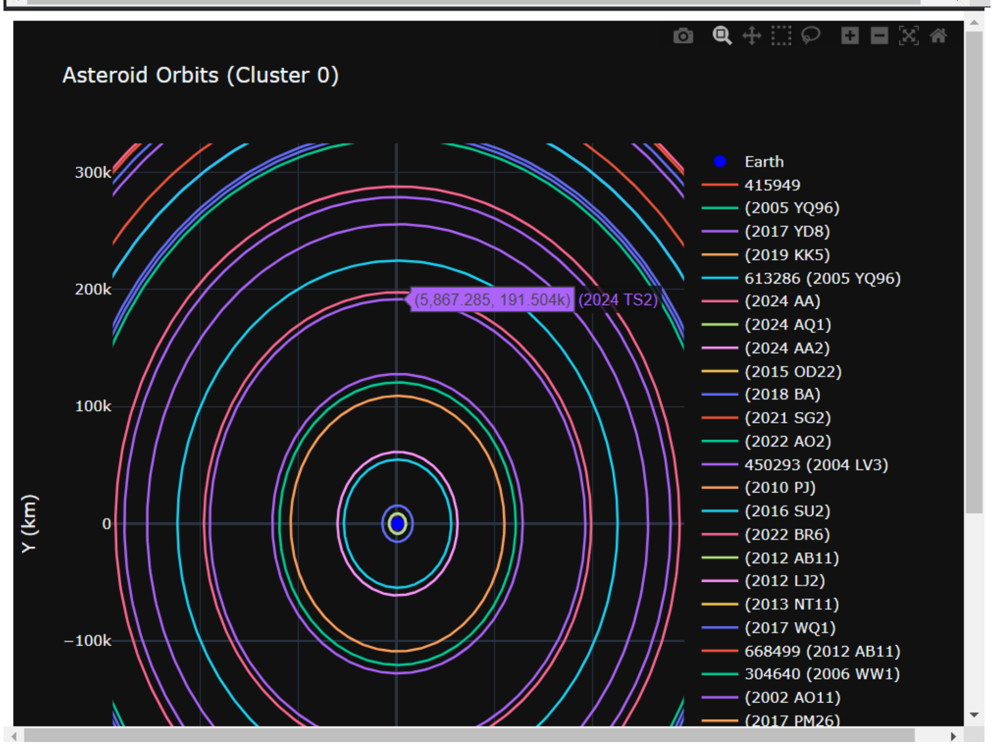
Image 6 shows Feature importance bar plot from the Random Forest classifier, showing the relative contributions of various capabilities to the model's predictions.

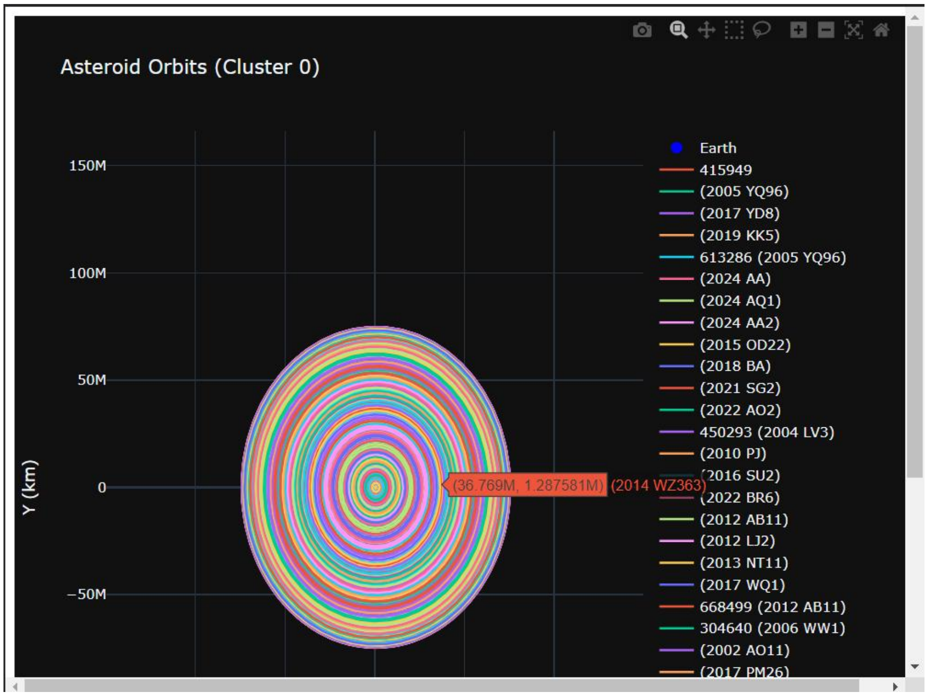
Orbit Simulation

A key aspect of the solution changed into the simulation of asteroid orbits relative to Earth. The simulation become carried out the usage of orbital factors, which have been converted into Cartesian coordinates. Python's Matplotlib library became used to create 2D and three-D visualizations of asteroid trajectories. These visualizations depicted the relative positions of asteroids over the years, highlighting ability collision situations.

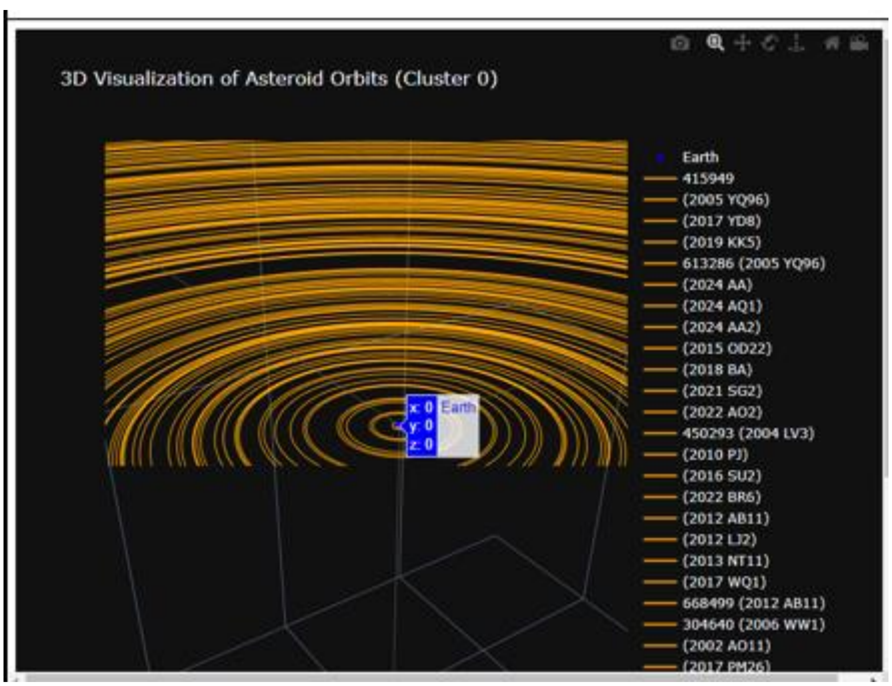
The simulation become extended to an interactive internet-based totally application the usage of Babylon.Js. The web software represented the solar gadget with practical models of planets and dynamically located asteroids. Each asteroid's trajectory became lively, permitting users to discover their motion relative to Earth. Hover results supplied additional statistics about each asteroid, which includes its size, speed, and omit distance. The simulation also included a clustering overlay, distinguishing excessive-hazard and occasional-hazard asteroids the usage of shade codes.

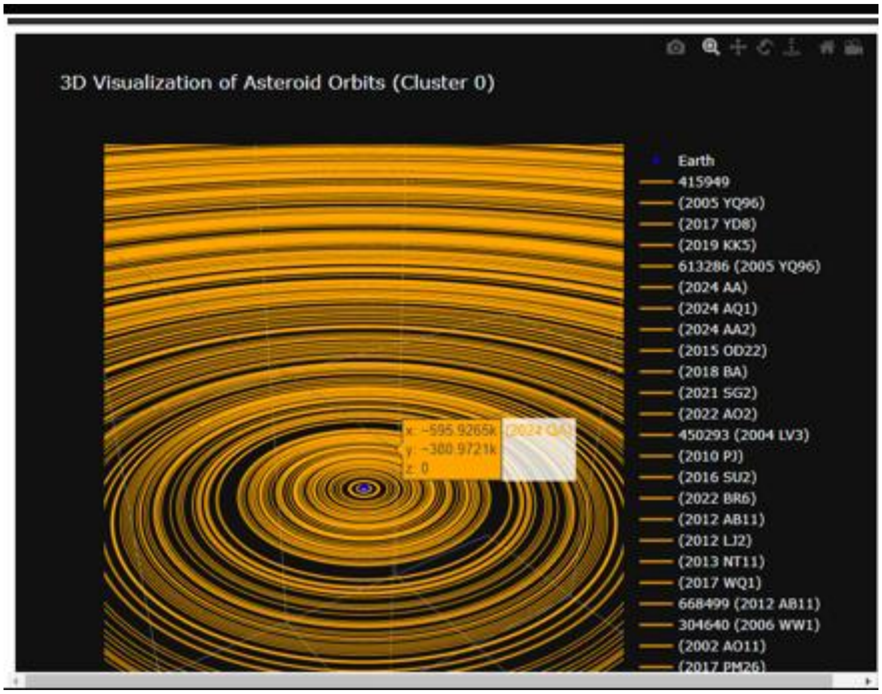

Asteroid Orbits (Cluster 0)
Y (km)
2M
1M
0
−1M
(−464.682k, 297.0683k) (2024 NG4)
Earth
415949
(2005 YQ96)
(2017 YD8)
(2019 KK5)
613286 (2005 YQ96)
(2024 AA)
(2024 AQ1)
(2024 AA2)
(2015 OD22)
(2018 BA)
(2021 SG2)
(2022 AO2)
450293 (2004 LV3)
(2010 PJ)
(2016 SU2)
(2022 BR6)
(2012 AB11)
(2012 LJ2)
(2013 NT11)
(2017 WQ1)
668499 (2012 AB11)
304640 (2006 WW1)
(2002 AO11)


Asteroid Orbits (Cluster 0)
Y (km)
300k
200k
100k
0
−100k
(5,867.285, 191.504k) (2024 TS2)
Earth
415949
(2005 YQ96)
(2017 YD8)
(2019 KK5)
613286 (2005 YQ96)
(2024 AA)
(2024 AQ1)
(2024 AA2)
(2015 OD22)
(2018 BA)
(2021 SG2)
(2022 AO2)
450293 (2004 LV3)
(2010 PJ)
(2016 SU2)
(2022 BR6)
(2012 AB11)
(2012 LJ2)
(2013 NT11)
(2017 WQ1)
668499 (2012 AB11)
304640 (2006 WW1)
(2002 AO11)

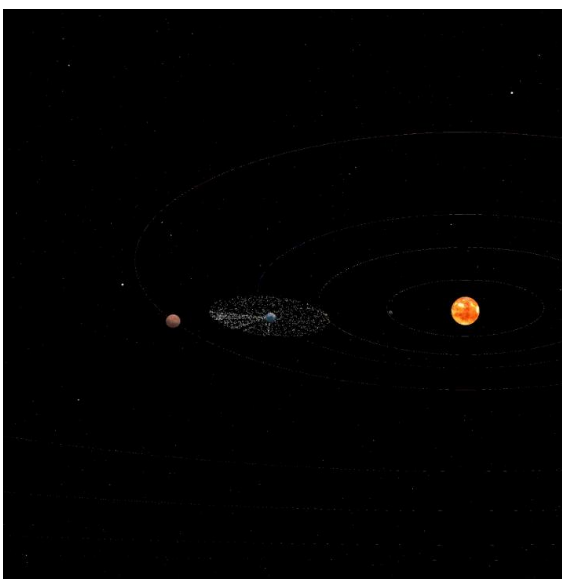


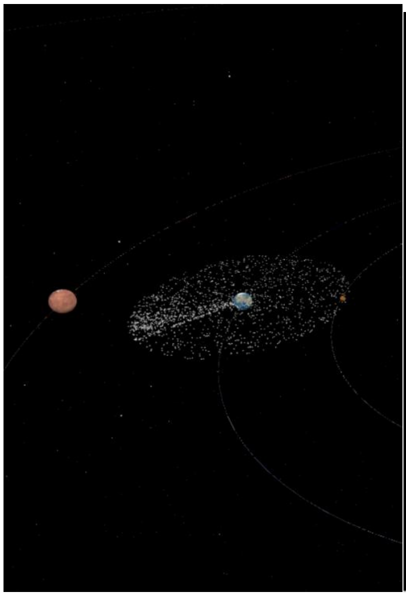
Images above are 2D visualization of asteroid orbits (Cluster zero) relative to Earth, displaying the paths of high-danger asteroids. Each colored circle is represented as the orbit of the asteroids and if you hover over the orbit you can see which asteroid it belongs to and their properties.

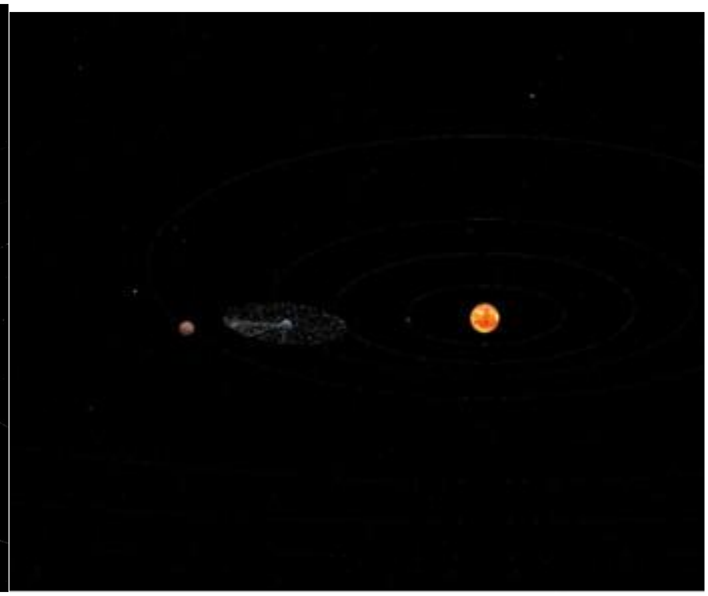


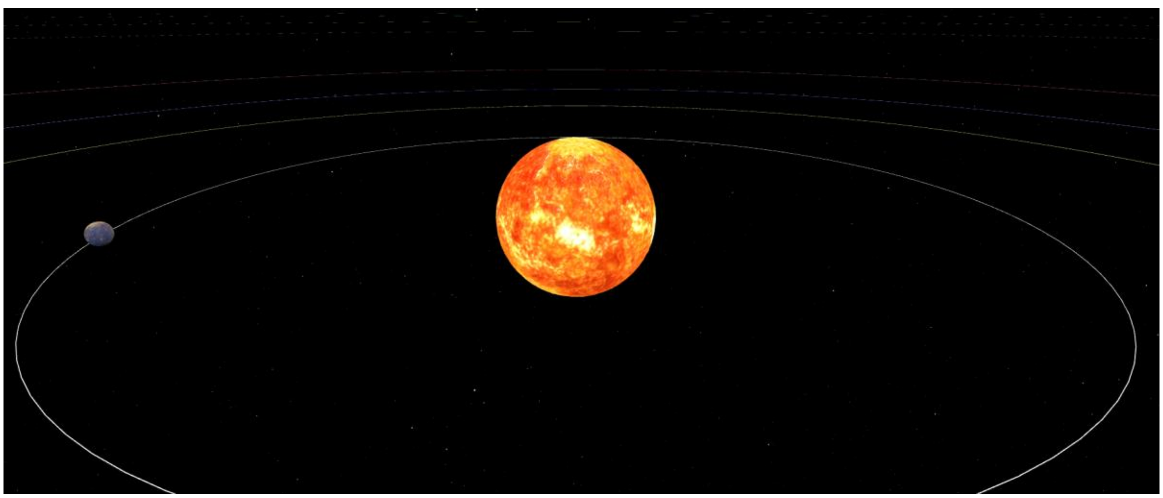


Images above represents the 3D visualization of asteroid orbits (Cluster zero), supplying a spatial angle of asteroid trajectories on the subject of Earth.









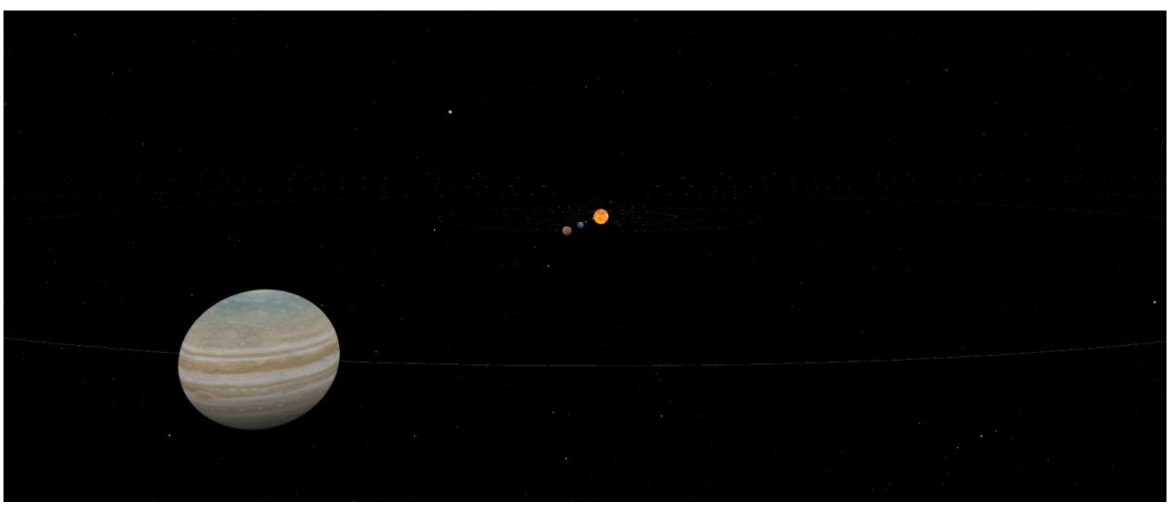


Image 10 is a wide three-D simulation of the sun device and asteroid trajectories, depicting the spatial relationship among planets and asteroids. I used React and BABYLON Library to make this simulation from scratch and in the link below you can see the solar system simulation without the asteroids, it gives you a better understanding of the simulation.

Link:
https://drive.google.com/file/d/1A0dxV2AwHy5VQlEKkO_pt2wSRVj91UyO/view?usp=sharing

Visualization and Insights

The visualization phase aimed to make the effects available and comprehensible. Key plots included:

K-Means Clustering Scatter Plot: This plot displayed the distribution of asteroids across speed and altitude, shade-coded by using cluster. It highlighted the separation between high-chance and occasional-hazard organizations.

Boxplots by using Cluster: These plots confirmed the distributions of man or woman features, together with speed, altitude, and strength, within every cluster. They provided statistical validation of the clustering results.

Feature Importance Bar Plot: This plot ranked the importance of functions inside the Random Forest classifier, supplying insights into the important thing predictors of collision hazard.

2D and 3D Orbit Visualizations: These visualizations depicted the relative positions of asteroids over the years, emphasizing their proximity to Earth.

The visualizations were instrumental in validating the model's predictions and communicating the consequences to stakeholders. They additionally served as an educational tool, illustrating the dynamics of asteroid movement and the factors influencing collision risks.

Integration and Deployment

The answer changed into implemented as a modular gadget, integrating facts processing, machine getting to know, and visualization additives. The trained fashions and processed datasets were serialized using joblib, ensuring reproducibility. The net utility turned into deployed as a standalone tool, reachable to customers for exploring asteroid trajectories and collision risks.

The deployment leveraged light-weight and scalable frameworks, making sure clean overall performance across gadgets. The system become designed to deal with updates to the datasets, permitting actual-time predictions as new data turns into available. This functionality is especially vital for planetary defense, where well timed and accurate information can manual mitigation efforts.

Future Directions

While the answer achieved its number one objectives, several regions for improvement remain. Incorporating extra datasets, which include the ones taking pictures gravitational interactions and long-term orbital modifications, could beautify the model's accuracy. Ensemble models combining more than one classifiers may additionally further improve predictive overall performance. Extending the simulation to consist of actual-time updates and API integrations with NASA's systems may want to make the tool even extra valuable.

In conclusion, this assignment demonstrates a comprehensive technique to predicting asteroid collisions and visualizing their dynamics. By combining clustering, type, and simulation, it gives a powerful framework for know-how and mitigating the dangers posed via potentially hazardous asteroids. The interactive internet application similarly complements its accessibility, making it a valuable tool for both researchers and the general public.

PERFORMANCE EVALUATIONS

The performance assessment section of this task changed into vital to assessing the accuracy, reliability, and performance of the asteroid collision prediction model and visualization gear. This section outlines the methodologies used for assessment, the outcomes received, and the technologies hired at some stage in the challenge.

Machine Learning Model Performance

The Random Forest Classifier become the primary supervised getting to know algorithm used on this task for predicting asteroid collision risks. Its overall performance became evaluated on a couple of metrics, together with accuracy, precision, don't forget, F1-score, and function significance. The following summarizes the key assessment effects:

Accuracy: The model done an accuracy of 65% at the test dataset. This shows that the version effectively categorised 65% of the asteroid times as both excessive-hazard or low-chance.
Precision: The precision score for the excessive-threat magnificence become 0.Seventy two, demonstrating that seventy two% of the asteroids expected as excessive-risk have been indeed excessive-threat.
Recall: The remember for the high-danger magnificence became 0.Sixty eight, that means that 68% of the real high-chance asteroids had been successfully diagnosed by means of the model.

F1-Score: The F1-score, which balances precision and recall, become zero.70 for the high-danger elegance, indicating a fairly balanced overall performance.
The function significance analysis found out that longitude, pace additives (vx and vy), and altitude had been the most influential features inside the class. These findings align with the predicted bodily standards governing asteroid trajectories, validating the relevance of the chosen capabilities.

Clustering Performance

The K-Means clustering algorithm changed into used to organization asteroids into high-risk and occasional-danger clusters primarily based on ancient information. The fine of the clustering became evaluated the usage of visualizations and statistical comparisons between clusters. Scatter plots and boxplots confirmed clean separations between high-threat and low-threat agencies throughout more than one capabilities, consisting of velocity, altitude, and radiated energy.

Cluster centroids had been analyzed to apprehend the important thing characteristics of each institution. The high-danger cluster exhibited better velocities and altitudes on common, constant with the bodily elements contributing to collision dangers. The clustering outcomes supplied a robust foundation for labeling the data, allowing the following supervised studying step.

Simulation and Visualization Performance

The 2D and three-D simulations of asteroid orbits had been evaluated based totally on their accuracy, usability, and computational performance. The simulation framework as it should be modeled the trajectories of asteroids the usage of orbital factors, providing reliable visualizations in their relative positions to Earth over the years. The following performance elements have been noted:

Accuracy: The simulations matched the predicted orbital paths of asteroids based totally on their enter parameters. This was proven with the aid of comparing the simulated trajectories with facts from NASA's publicly to be had tools.

Usability: The net-primarily based simulation, developed the use of Babylon.Js, provided an intuitive interface for exploring asteroid movement. Hover results and clustering overlays stronger user engagement and understanding.

Efficiency: The visualization equipment performed efficiently, with minimal lag or performance problems, even when rendering multiple asteroid trajectories simultaneously.

Tools and Environments

The implementation and evaluation of the project relied on a combination of software equipment, programming languages, and hardware assets. These technologies played a essential position in ensuring the achievement of the undertaking:

Programming Languages:

Python: Used for records processing, clustering, system studying, and static visualizations. Key libraries protected Pandas, NumPy, scikit-research, Matplotlib, and Seaborn.

JavaScript: Used for developing the interactive web-based totally simulation the use of the Babylon.Js framework.

Machine Learning Tools:

scikit-examine: Provided implementations of the Random Forest Classifier, K-Means clustering, and facts preprocessing utilities which include StandardScaler.

Visualization Tools:

- Matplotlib and Seaborn: Used for growing 2D scatter plots, boxplots, and function significance bar plots.
- Babylon.Js: Enabled the development of a 3-D net-primarily based simulation, providing an interesting visualization of asteroid orbits.Development Environment:
- IDE: PyCharm and Visual Studio Code were used for Python and JavaScript development, respectively.
- Web Development: Node.Js and npm have been used for managing dependencies and constructing the net utility.

Data Sources:

NASA's CNEOS Fireball database and asteroid orbital records had been the number one statistics resources for the mission.

CONCLUSIONS

This venture addresses the crucial task of predicting probably dangerous asteroid (PHA) collisions with Earth via integrating facts processing, machine mastering, and visualization techniques. Several key issues emerged all through the challenge, which include the complexity of asteroid trajectory prediction, the confined availability of unified datasets, and the want for correct simulations. These demanding situations have been mitigated via a established approach that blended clustering, supervised studying, and interactive visualizations.

One of the principal advantages of this solution method is its capacity to combine historical and modern-day datasets, permitting a unified analysis of asteroid collision risks. The use of K-Means clustering supplied precious insights into the traits of high-chance asteroids, at the same time as the Random Forest Classifier prolonged these insights to new facts factors with an affordable accuracy of sixty five%. The inclusion of feature significance evaluation ensured the interpretability of the model, making it viable to become aware of key predictors of collision risk. Additionally, the interactive internet-based totally simulation improved the accessibility and value of the consequences, imparting an intuitive platform for exploring asteroid dynamics.

However, the assignment also has sure limitations. The accuracy of sixty five%, at the same time as reasonable, leaves room for improvement. The clustering and type fashions depended on a enormously small set of features, and incorporating additional parameters, which include gravitational interactions and lengthy-term orbital adjustments, could improve the predictions. The simulation framework, even though efficient, can be in addition more suitable by means of integrating actual-time information updates and outside APIs from groups like NASA.

The contributions of this undertaking are large. It affords a comprehensive framework for knowledge and mitigating asteroid collision dangers, bridging the gap between astrophysics and computational technology. The modular design guarantees scalability, even as using open information and tools makes the solution accessible to a much wider target audience. The assignment also serves as a precious instructional aid, demonstrating the application of device gaining knowledge of and visualization in solving actual-international troubles.

Future work may want to focus on numerous areas of development. Enhancing the device mastering fashions by way of incorporating ensemble strategies and expanding the feature set could in all likelihood improve predictive accuracy. Real-time updates and integration with outside APIs ought to make the simulation device more dynamic and informative. Additionally, extending the scope of the mission to include different celestial threats, which includes comets or area debris, could similarly increase its effect.
In end, this venture demonstrates a hit utility of multidisciplinary techniques to cope with a vital planetary defense challenge. While there are opportunities for improvement, the answer offers a strong foundation for destiny improvements in asteroid tracking and collision danger prediction.

REFERENCES

[1] NASA Center for Near Earth Object Studies (CNEOS), "Close Approaches," [Online]. Available: https://cneos.jpl.nasa.gov/ca/. [Accessed: Jan. 11, 2025].

[2] NASA Center for Near Earth Object Studies (CNEOS), "Introduction to Close Approaches," [Online]. Available: https://cneos.jpl.nasa.gov/ca/intro.html. [Accessed: Jan. 11, 2025].

[3] NASA Center for Near Earth Object Studies (CNEOS), "Fireball and Bolide Data," [Online]. Available: https://cneos.jpl.nasa.gov/fireballs/. [Accessed: Jan. 11, 2025].

[4] NASA Center for Near Earth Object Studies (CNEOS), "Planetary Defense Coordination Office - Potential Impacts," [Online]. Available: https://cneos.jpl.nasa.gov/pi/. [Accessed: Jan. 11, 2025].

[5] École Polytechnique Fédérale de Lausanne (EPFL), "Space Mission Design and Operations," edX, [Online]. Available: https://www.edx.org/learn/space/ecole-polytechnique-federale-de-lausanne-space-mission-design-and-operations?index=product&queryId=fb43d6eb6faa3dabd5b17c6661c5ba7c&position=1. [Accessed: Jan. 11, 2025].

[6] Udemy, "Python for Time Series Data Analysis," [Online]. Available: https://www.udemy.com/course/python-for-time-series-data-analysis/?couponCode=ACCAGE0923. [Accessed: Jan. 11, 2025].

[7] C. R. Chapman, D. Morrison, and B. Zellner, "Asteroids: Past and Future," Science, vol. 211, no. 4474, pp. 1211–1214, Mar. 1981, doi: 10.1126/science.211.4474.1211.

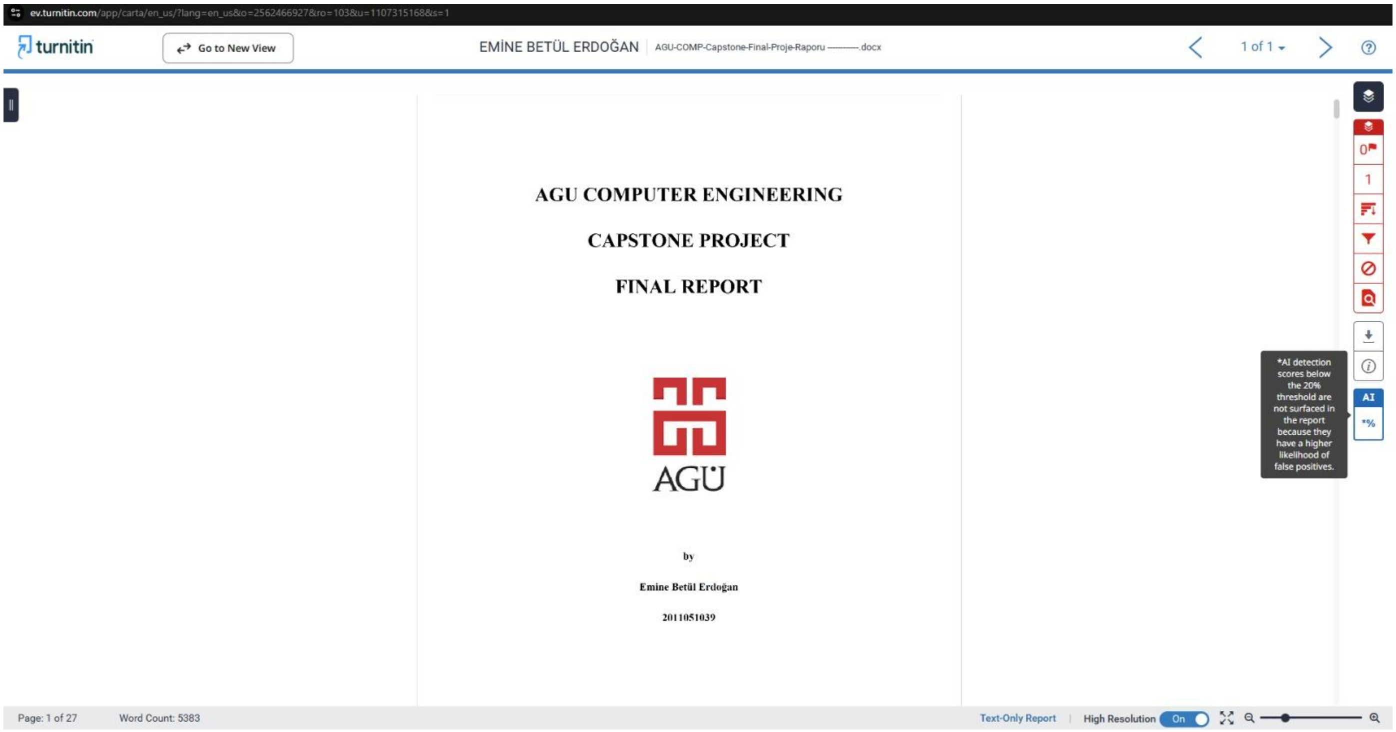
[8] A. W. Harris and G. D’Abramo, "The Population of Near-Earth Asteroids," Icarus, vol. 257, pp. 302–312, May 2015, doi: 10.1016/j.icarus.2015.05.004.

[9] P. G. Brown et al., "The Chelyabinsk Airburst: Implications for the Hazard Posed by Small NEOs," Nature, vol. 503, pp. 238–241, Nov. 2013, doi: 10.1038/nature12741.

[10] R. Jedicke, J. Granvik, and M. Micheli, "The Near-Earth Object Population," Astronomy and Astrophysics Review, vol. 25, no. 1, pp. 1–19, Dec. 2017, doi: 10.1007/s00159-017-0104-9.

APPENDIX

Author	Title	Submission ID	Uploaded	Viewed	Similarity	Flags	Options
EMİNE BETÜL ERDOĞAN	AGU-COMP-Capstone-Fi...	2562466927	January 11, 2025		1%	--	•••


ev.turnitin.com/app/carta/en_us/?lang=en_us&o=2562466927&ro=103&u=1107315168&s=1
turnitin
Go to New View
EMİNE BETÜL ERDOĞAN
AGU-COMP-Capstone-Final-Proje-Raporu ———.docx
1 of 1
AGU COMPUTER ENGINEERING
CAPSTONE PROJECT
FINAL REPORT
AGÜ
by
Emine Betül Erdoğan
2011051039
*AI detection scores below the 20% threshold are not surfaced in the report because they have a higher likelihood of false positives.
AI
*%
Page: 1 of 27
Word Count: 5383
Text-Only Report
High Resolution
On