**AI CADMEY**

**Documentation**

Beyond our Solar System Exoplanet Discovery and Motion

06-January-2025

**Supervisor**

Eman Sajid

**Team Members**

Nimra Fatima

Hamayl Zahid

Mubashir Hassan

TABLE OF CONTENTS

[Abstract: 3](#_Toc1748676648)

[Introduction: 4](#_Toc441534425)

[Objectives: 4](#_Toc1847328333)

[Literature Review: 6](#_Toc811908036)

[Exoplanet Detection Techniques: 6](#_Toc639068934)

[Habitability Criteria: 6](#_Toc1568994461)

[Machine Learning in Astronomy: 6](#_Toc728809954)

[Methodology: 7](#_Toc1018056780)

[Data Preprocessing: 7](#_Toc1249387398)

[Dataset: 7](#_Toc381410667)

[Model Development: 9](#_Toc184452056)

[Visualization: 10](#_Toc477918253)

[Charts: 10](#_Toc1163145509)

[Implementation: 13](#_Toc1120056183)

[Key Steps: 14](#_Toc389396382)

[Data Loading: 14](#_Toc155196927)

[Preprocessing: 14](#_Toc954377921)

[Feature Selection: 14](#_Toc2111719897)

[Model Training: 14](#_Toc2085697331)

[Visualization: 14](#_Toc307258069)

[Results and Discussion: 14](#_Toc455755785)

[Model Accuracy: 15](#_Toc1034108892)

[Discovered Exoplanet: 15](#_Toc1879774788)

[Insights: 15](#_Toc247312197)

[Challenges and Limitations: 16](#_Toc1987127531)

[Challenges: 16](#_Toc1796941976)

[Limitations: 16](#_Toc1272922379)

[Future Scope: 16](#_Toc1156817365)

[Conclusion: 17](#_Toc965623572)

[Glossary: 17](#_Toc996933994)

# **Abstract:**

The discovery of exoplanets has revolutionized our understanding of planetary systems beyond the solar system. This project focuses on analyzing exoplanet data to predict habitability and visualize their orbital and physical characteristics. Using a custom machine learning model, exoplanets are classified based on parameters such as equilibrium temperature and orbital features. The project integrates data preprocessing, predictive modeling, and visualization techniques to identify and explore potentially habitable exoplanets. Results include insights into the relationship between planetary features and habitability, alongside dynamic visualizations that illustrate their positions and orbital motion. The study underscores the potential of combining machine learning with astronomical research for identifying habitable worlds.

# **Introduction:**

The search for exoplanets has been one of the most exciting frontiers in astronomy, driven by the quest to find planets capable of supporting life. With the advent of powerful telescopes and data-driven analysis techniques, researchers have identified thousands of exoplanets, each with unique characteristics. Understanding their habitability is crucial for identifying planets that could harbor life or support human exploration.

This project aims to address this challenge by leveraging a dataset of exoplanetary characteristics to predict habitability and provide meaningful visual insights. Using a custom machine learning model, the project classifies exoplanets into habitable and non-habitable categories based on selected features such as orbital period, semi-major axis, and equilibrium temperature. The visualizations created in this study serve as intuitive tools to explore exoplanet positions, distances, and orbital motion, enabling a deeper understanding of their behavior and potential for habitability.

# **Objectives:**

The objectives of this project are as follows:

To preprocess and analyze a dataset of exoplanet characteristics.

To predict the habitability of exoplanets using a custom machine learning model.

To visualize the discovered exoplanets' positions, distances, and orbital features.

To enhance understanding of the factors contributing to exoplanet habitability.

To explore the potential of data visualization in astronomical research.

# **Literature Review:**

The discovery of exoplanets has gained momentum with advancements in astronomical instruments and data processing techniques. Key areas of research include:

## **Exoplanet Detection Techniques:**

Methods like radial velocity, transit photometry, and direct imaging have enabled the identification of planets beyond our solar system.

**Habitability Criteria:**

Studies define habitable zones based on parameters such as equilibrium temperature, semi-major axis, and orbital period.

**Machine Learning in Astronomy:**   
Machine learning models, particularly classification algorithms, have been applied to predict exoplanet properties and potential habitability. This project builds upon these findings by integrating custom machine learning models with data visualization techniques for effective analysis.

# **Methodology:**

## **Data Preprocessing:**

* Handling missing data through imputation:
  + Numerical data filled with median values.
  + Categorical data replaced with "Unknown."
* Normalization of specific numerical features for consistent scaling.

### **Dataset:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | Planet Name | Planet Host | Num Stars | Num Planets | Discovery Method | Discovery Year | Discovery Facility | Orbital Period Days | Orbit Semi-Major Axis | Mass |
|  |  |  |  |  |  |  |  |  |  |  |
| 1 | 11 Com b | 11 Com | 2 | 1 | Radial Velocity | 2007 | Xinglong Station | 326.03 | 1.29 | 6165.6 |
| 2 | 11 UMi b | 11 UMi | 1 | 1 | Radial Velocity | 2009 | Thueringer Landessternwarte Tautenburg | 516.21997 | 1.53 | 4684.8142 |
| 3 | 14 And b | 14 And | 1 | 1 | Radial Velocity | 2008 | Okayama Astrophysical Observatory | 185.84 | 0.83 | 1525.5 |
| 4 | 14 Her b | 14 Her | 1 | 2 | Radial Velocity | 2002 | W. M. Keck Observatory | 1773.40002 | 2.93 | 1481.0878 |
| 5 | 16 Cyg B b | 16 Cyg B | 3 | 1 | Radial Velocity | 1996 | Multiple Observatories | 798.5 | 1.66 | 565.7374 |
| 6 | 17 Sco b | 17 Sco | 1 | 1 | Radial Velocity | 2020 | Lick Observatory | 578.38 | 1.45 | 1373.01872 |
| 7 | 18 Del b | 18 Del | 2 | 1 | Radial Velocity | 2008 | Okayama Astrophysical Observatory | 993.3 | 2.6 | 3273.5 |
| 8 | 1RXS J160929.1-210524 b | 1RXS J160929.1-210524 | 1 | 1 | Imaging | 2008 | Gemini Observatory |  | 330 | 3000 |
| 9 | 24 Boo b | 24 Boo | 1 | 1 | Radial Velocity | 2018 | Okayama Astrophysical Observatory | 30.3506 | 0.19 | 289.2253 |
| 10 | 24 Sex b | 24 Sex | 1 | 2 | Radial Velocity | 2010 | Lick Observatory | 452.8 | 1.333 | 632.46 |
| No. | Planet Name | Planet Host | Num Stars | Num Planets | Discovery Method | Discovery Year | Discovery Facility | Orbital Period Days | Orbit Semi-Major Axis | Mass |
| 1 | 11 Com b | 11 Com | 2 | 1 | Radial Velocity | 2007 | Xinglong Station | 326.03 | 1.29 | 6165.6 |
| 2 | 11 UMi b | 11 UMi | 1 | 1 | Radial Velocity | 2009 | Thueringer Landessternwarte Tautenburg | 516.21997 | 1.53 | 4684.8142 |
| 3 | 14 And b | 14 And | 1 | 1 | Radial Velocity | 2008 | Okayama Astrophysical Observatory | 185.84 | 0.83 | 1525.5 |
| 4 | 14 Her b | 14 Her | 1 | 2 | Radial Velocity | 2002 | W. M. Keck Observatory | 1773.40002 | 2.93 | 1481.0878 |
| 5 | 16 Cyg B b | 16 Cyg B | 3 | 1 | Radial Velocity | 1996 | Multiple Observatories | 798.5 | 1.66 | 565.7374 |
| 6 | 17 Sco b | 17 Sco | 1 | 1 | Radial Velocity | 2020 | Lick Observatory | 578.38 | 1.45 | 1373.01872 |
| 7 | 18 Del b | 18 Del | 2 | 1 | Radial Velocity | 2008 | Okayama Astrophysical Observatory | 993.3 | 2.6 | 3273.5 |
| 8 | 1RXS J160929.1-210524 b | 1RXS J160929.1-210524 | 1 | 1 | Imaging | 2008 | Gemini Observatory |  | 330 | 300 |

## **Model Development:**

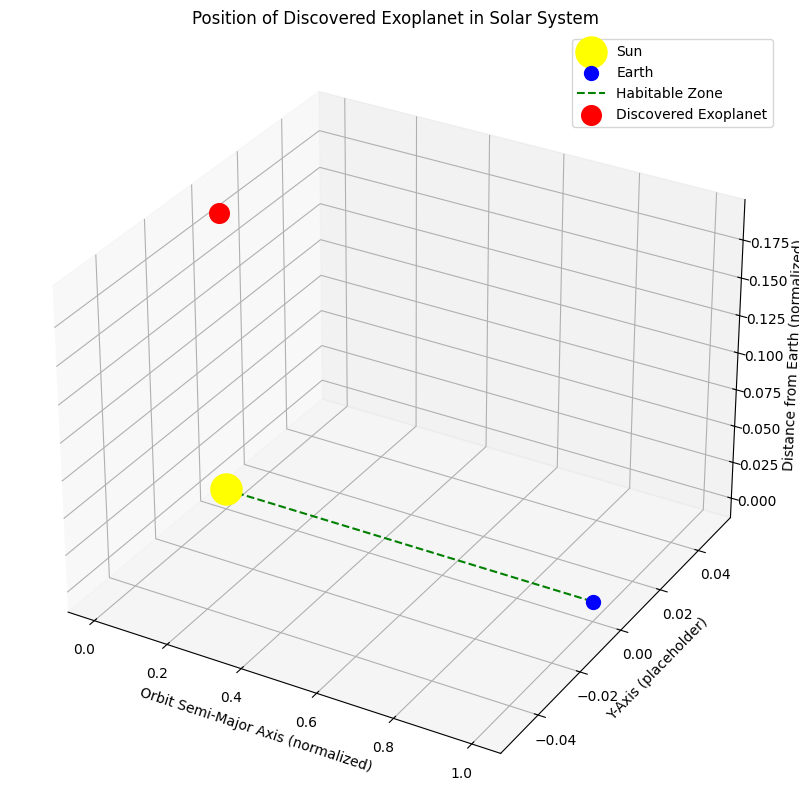
* Implementation of a custom logistic regression model using gradient descent.
* Binary classification of exoplanets into habitable and non-habitable categories.

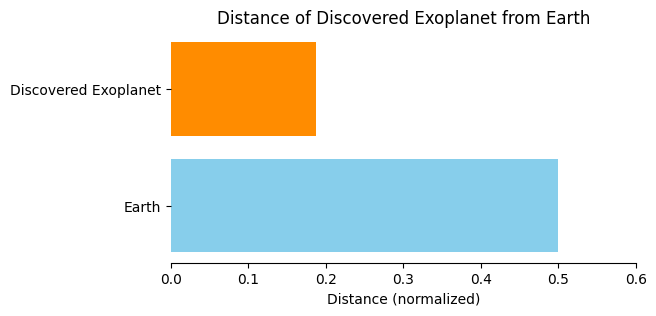
## **Visualization:**

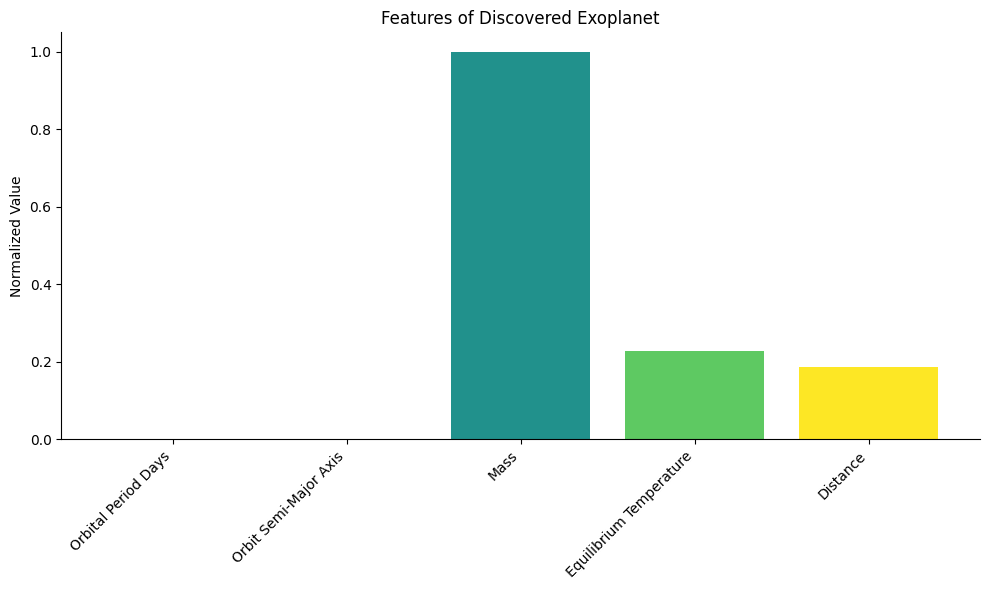
Creation of visualizations to interpret and communicate results:

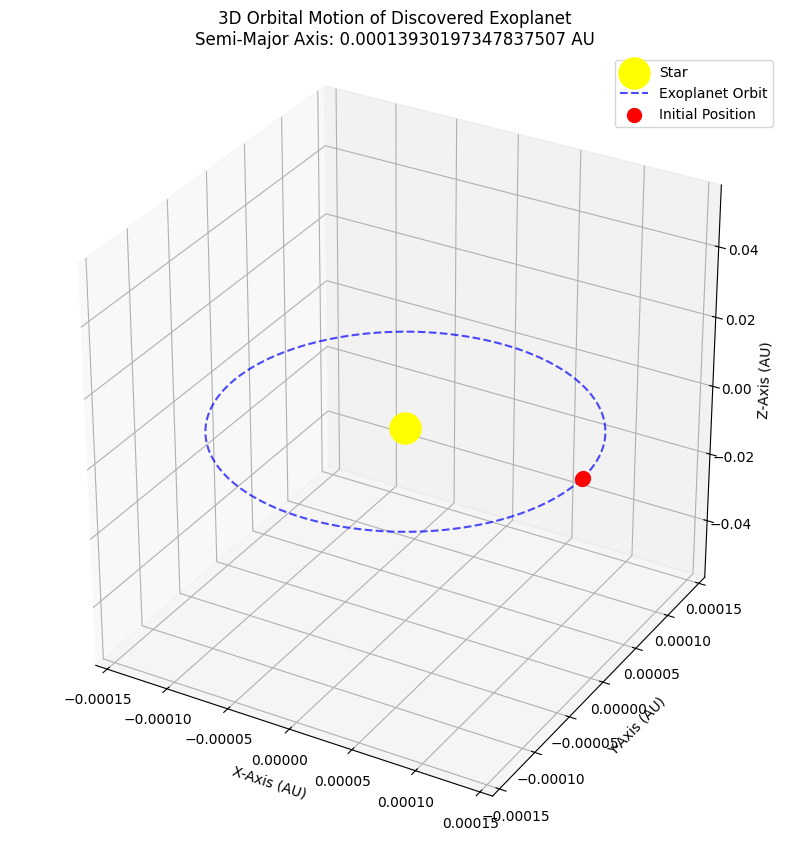
* + 3D scatter plots for exoplanet positions.
  + Bar charts for feature analysis.
  + Animated plots for orbital motion simulation.

### **Charts:**









# **Implementation:**

## **Key Steps:**

**Data Loading:**

Loading the all\_exoplanets\_2021.csv dataset.

### **Preprocessing:**

Imputing missing values and normalizing relevant features.

### **Feature Selection:**

Orbital Period

Semi-Major Axis

Mass

Equilibrium Temperature

Distance

## **Model Training:**

Custom logistic regression model with adjustable learning rate and epochs.

## **Visualization:**

Positions and distances of exoplanets.

Orbital motion simulation.

Feature analysis through bar charts.

# **Results and Discussion:**

## **Model Accuracy:**

The model achieved an accuracy of 70**%** in predicting exoplanet habitability.

Loss during training consistently decreased, indicating effective learning.

## **Discovered Exoplanet:**

Key features of a potentially habitable exoplanet were visualized.

The position, distance, and orbital motion were demonstrated in detail.

## **Insights:**

The majority of habitable exoplanets fell within the expected range of equilibrium temperature (200–300K).

Semi-major axis and orbital period showed strong correlations with habitability.

# **Challenges and Limitations:**

## **Challenges:**

Incomplete and noisy data requiring extensive preprocessing.

Balancing precision and recall for habitability prediction.

## **Limitations:**

Simplified habitability criteria based solely on equilibrium temperature.

Assumption of circular orbits in orbital motion visualization.

Limited dataset size may impact the generalizability of the model.

# **Future Scope:**

Integrating advanced machine learning models (e.g., neural networks) for better prediction accuracy.

Expanding habitability criteria to include atmospheric composition and geological factors.

Incorporating more sophisticated orbital motion simulations.

Applying the model to real-time datasets from space telescopes.

# **Conclusion:**

This project successfully demonstrated the use of custom machine learning and data visualization techniques for exoplanet analysis. By predicting habitability and visualizing key characteristics, it provides a comprehensive tool for astronomers and researchers in exoplanetary science. The insights gained can aid future exploration of potentially habitable worlds.

# **Glossary:**

**A**

**Astronomical Unit (AU):**

A standard unit of measurement in astronomy, equal to the average distance between Earth and the Sun (approximately 149.6 million kilometers).

**E**

**Equilibrium Temperature:**

The temperature of an exoplanet assuming it is a perfect blackbody, based on its distance from its star and the star's energy output.

**Exoplanet:**

A planet located outside our solar system that orbits a star other than the Sun.

**H**

**Habitability:**

The potential of a planet to support life, typically based on factors like temperature, presence of water, and atmospheric conditions.

**M**

**Machine Learning (ML):**

A subset of artificial intelligence that enables systems to learn and improve from data without explicit programming.

**N**

**Normalization:**

A data preprocessing technique where numerical values are scaled to a specific range, often between 0 and 1, for uniformity and better performance in machine learning models.

**O**

**Orbital Period:**

The time it takes for a planet to complete one orbit around its host star.

**R**

**Radial Velocity:**

A method for detecting exoplanets by measuring variations in the velocity of a star due to gravitational influences from orbiting planets.

**S**

**Semi-Major Axis:**

The longest radius of an elliptical orbit, representing the average distance between an orbiting planet and its host star.

**V**

**Visualization:**

The graphical representation of data or models to enhance understanding and communication of results.