Habitable Exoplanet Explorer

Project Team

Muhammad Ahmed Raza 19P-0070

Session 2019-2024

Supervised by

Dr. Omar Usman Khan

Department of Computer Science

National University of Computer and Emerging Sciences
Peshawar, Pakistan

December, 2024

Student's Declaration

I declare that this project titled "*Habitable Exoplanet Explorer*", submitted as requirement for the award of degree of Bachelors in Computer Science, does not contain any material previously submitted for a degree in any university; and that to the best of my knowledge, it does not contain any materials previously published or written by another person except where due reference is made in the text.

I understand that the management of Department of Computer Science, National University of Computer and Emerging Sciences, has a zero tolerance policy towards plagiarism. Therefore, I, as an author of the above-mentioned thesis, solemnly declare that no portion of my thesis has been plagiarized and any material used in the thesis from other sources is properly referenced.

I further understand that if I am found guilty of any form of plagiarism in the thesis work even after graduation, the University reserves the right to revoke my BS degree.

Muhammad Ahmed Raza

Signature: ________________

____

Verified by Plagiarism Cell Officer

Dated:

Certificate of Approval

The Department of Computer Science, National University of Computer and Emerging Sciences, accepts this thesis titled *Habitable Exoplanet Explorer*, submitted by Muhammad Ahmed Raza (19P-0070), in its current form, and it is satisfying the dissertation requirements for the award of Bachelors Degree in Computer Science.

Supervisor

Dr. Omar Usman Khan Signature: ________________

Mr. Haroon Zafar

FYP Coordinator
National University of Computer and Emerging Sciences, Peshawar

Mr. Fazl-e-Basit

HoD of Department of Computer Science
National University of Computer and Emerging Sciences

Acknowledgements

I would like to start by mentioning Allah (SWT), the Most Gracious and the Most Merciful. Allah (SWT) granted me the strength to finish this project and I am truly grateful for all the opportunities, challenges, and the blessings that came along the way. Many people directly or indirectly played a role in this, specially our parents, teachers who motivated me. I'd like to express my earnest appreciation to Dr. Omar Usman for administering this project.

Muhammad Ahmed Raza

Abstract

In this study, we explored and analyzed the exoplanet data to better understand how exoplanets might be habitable. What features, and circumstances lead to a better habitability of a planet. What could it mean for a potential extraterrestrial life. In our research, we acquired and analyzed data from publicly available NASA's Kepler and TESS missions. In order to predict the probability of habitability from core stellar and exoplanet features, we used Python for data processing and machine learning techniques to develop models. We identify promising candidates for additional astrobiology investigation and stress the importance of habitability assessment.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1. **Purpose of the Investigation:** The purpose of this investigation is to explore the exoplanets data for potential habitability through thorough data analysis and statistical modeling.
2. **Problem Being Investigated:** The problem being investigated is the challenge of processing and interpreting vast amounts of exoplanet data to determine habitability.
3. **Background:** Convincing the public that space exploration is not only possible, but that, life is possible beyond planet Earth is vital in the context of astrobiology, as we stand to understand the habitability of exoplanets. Researchers have done groundwork for the discovery of exoplanets but there is an ill need for sophisticated tools to analyze and interpret this data.
4. **Thesis and General Approach:** We argue that using data analysis and machine learning we can create a system that will predict habitable exoplanets more accurately and efficiently. For this we will use techniques such as data preprocessing, machine learning model development, and validation.
5. **Criteria for Study's Success:** The success of this study will be determined by the system's ability to predict habitability with a high degree of accuracy, as well as its usability and scalability for future research in space science.

Chapter 2

Review of Literature

2.1 Literature Review

Sr. no	Year	Key Points	Methodologies	Limitations
1	2020	The occurrence of Rocky Habitable zone Planets around Solar-like stars from Kepler Data	$\eta_{\oplus}$ for the conservative HZ is between $0.21^{+0.37}_{-0.21}$ and $0.51^{+0.88}_{-0.33}$ planets per star	Large uncertainties due to small sample size
2	2021	It focuses on how the core mass fraction (CMF) influences the thickness, composition, and mineralogy of the planet's crust, which in turn affects volatile cycling and habitability.	petrological modeling to simulate mantle decompression melting and crust production in planets with varying CMFs	Simplified model assumptions

Sr. no	Year	Key Points	Methodologies	Limitations
3	2015	the delivery of water to these planets, with a focus on the habitability of Earth.	the use of N-body simulations and Smoothed Particle Hydrodynamics (SPH) simulations to model planet formation and collisions.	Simplified collisions and initial conditions
4	2009	the possibility of planet formation in the habitable zone (HZ) of Alpha Centauri B, a binary star system.	numerical simulations to model the dynamics of planetesimals in the binary system. It considers the combined effects of the companion star's perturbations and gas drag on planetesimal accretion.	simulations are simplified and do not include all possible mechanisms that could influence planet formation in the HZ.
5	2007	the habitability of two planets (Gl 581c and Gl 581d) discovered around the M-type star Gliese 581. It focuses on estimating the boundaries of the habitable zone (HZ) around this star and the uncertainties involved in their determination.	The study provides simplified formulae for estimating the inner (lin) and outer (lout) edges of the habitable zone based on stellar luminosity (L) and effective temperature (Teff)	Uncertainties in HZ boundaries and atmospheric properties

Sr. no	Year	Key Points	Methodologies	Limitations
6	2007	This paper investigates the formation of habitable planets in binary star systems, where the primary star hosts a Jupiter-like planet. It focuses on how the secondary star's motion affects the formation and water content of Earth-like planets in the habitable zone.	Stability criterion for giant planets	Assumes terrestrial planet formation similar to single stars
7	2022	This study focuses on the formation of habitable planets around low-mass stars, particularly M dwarfs.	Water-ice snow line location	Simplified models, no gas-disk interactions

Table 2.1: Literature Review

2.2 Explanation

In Bryson et al. [2020], This paper demonstrates the occurrence rates for the rocky planets in the Habitable zones (HZs) of main-sequence dwarf stars based on Kepler data and Gaia-based stellar properties. The paper defines $\eta_{\oplus}$ as the HZ occurrence of planets with radii between 0.5 and 1.5 $R_{\oplus}$ orbiting stars with effective temperatures between 4800 K and 6300 K. It finds that $\eta_{\oplus}$ for the conservative HZ is between $0.21^{+0.37}_{-0.21}$ and $0.51^{+0.88}_{-0.33}$ planets per star, while the optimistic HZ occurrence is between $0.33^{+0.58}_{-0.33}$ and $0.73^{+0.88}_{-0.51}$ planets per star. The paper concludes that the occurrence rate for rocky planets per star depends upon the stellar effective temperature.

In Dyck et al. [2021], This study was done on the effect of core formation on the planet's surface and habitability of the rocky planets. It generally focuses on core mass fraction (CMF) which shows the influence on the thickness, composition and mineralogy of the planet's crust, which affects the habitability of the planet. The study uses petrological modeling to simulate mantle decompression melting, crust production in planets with different CMFs. The study concludes that the extent of core formation significantly impacts a planet's surface environment and habitability. Planets with large CMFs tend to have thin, anhydrous crusts, while those with smaller CMFs develop thicker crusts capable of sequestering volatiles and supporting a water cycle.

In Dvorak et al. [2015], This study was done on the formation of planets and delivery of water to these planets, particularly keeping in mind the habitability of the Earth. N-body simulations and Smoothed Particle Hydrodynamics (SPH) simulations are used to model planet formation and collisions to observe the role of collisions in the growth of planetesimals. Although, the study acknowledges that the outcome of planet formation simulations is highly dependent on the choice of initial conditions and realistic collision outcomes are complex and not fully understood.

In Thebault et al. [2009], This paper demonstrates the possibility of planet formation in the habitable zone (HZ) of Alpha Centauri B, a binary star system. Uses numerical simulations to model the dynamics of planetesimals, particularly the accretion of kilometer-sized planetesimals. The study acknowledges that its simulations are simplified and do not include all possible mechanisms that could influence planet formation in the HZ. It also notes the uncertainty in the initial conditions of the binary system, which could have evolved over time.

In Selsis et al. [2007], This paper assesses the habitability of two planets (Gl 581c and Gl 581d) discovered around the M-type star Gliese 581. It focuses on estimating the habitable zone (HZ) boundaries and uncertainties during its calculation, around the star.

The study provides simplified formulae for estimating the inner (l_{in}) and outer (l_{out}) edges of the habitable zone based on stellar luminosity (L) and effective temperature (T_{eff}):

$$l_{\text{in}} = \left(l_{\text{in},0} - a_{\text{in}}(T_{\text{eff}} - 5700) - b_{\text{in}}(T_{\text{eff}} - 5700)^2\right)\left(\frac{L}{L_\odot}\right)^{1/2}$$

$$l_{\text{out}} = \left(l_{\text{out},0} - a_{\text{out}}(T_{\text{eff}} - 5700) - b_{\text{out}}(T_{\text{eff}} - 5700)^2\right)\left(\frac{L}{L_\odot}\right)^{1/2}$$

where $l_{\text{in},0}$ and $l_{\text{out},0}$ are the inner and outer HZ limits for the Sun, and $a_{\text{in}}, b_{\text{in}}, a_{\text{out}}$, and b_{out} are constants. The study emphasizes uncertainties in determining the HZ boundaries due to multiple factors such as effects of clouds and the unknown properties of exoplanet atmosphere. The paper concludes that G1 581c is likely unhabitable due to its high stellar flux, while G1 581d, might be potentially be habitable, with conditions comparable to early Mars.

In Haghighipour and Raymond [2007], this paper emphasizes on the formation of habitable planets in binary star systems. It focuses on how the secondary star's motion affects the formation and water content of Earth-like planets in the habitable zone. The study uses the following formula to determine the stability of a giant planet's orbit in a binary system:

$$a_c = a_b\left[0.464 - 0.380\mu + (0.631 - 0.586\mu)\,e_b + (0.150 - 0.198\mu)\,e_b^2\right]$$

where a_c is the critical semimajor axis for stability, a_b and e_b are the semimajor axis and eccentricity of the binary companion, and μ is the mass ratio of the stars.

The study assumes that rocky planets in binary systems are formed similarly to that of single stars system and that giant planets have already been formed. it also simplifies the water delivery process by considering only protoplanetary objects. The formation and water content of these planets depend on the binary's separation and the eccentricity of the secondary star. Binaries with moderate to large perihelia and giant planets in low-eccentricity orbits are more favorable for habitable planet formation.

In Clement et al. [2022], this study focuses on the formation of habitable planets around low-mass stars, particularly M dwarfs. it emphasizes on the water delivery process through late-stage bombardment of icy asteroids and leftover planetesimals. The study uses the following formula to determine the stability of a giant planet's orbit in a binary system:

$$a_c = a_b \left[0.464 - 0.380\mu + (0.631 - 0.586\mu)\, e_b + (0.150 - 0.198\mu)\, e_b^2\right]$$

where a_c is the critical semimajor axis for stability, a_b and e_b are the semimajor axis and eccentricity of the binary companion, and μ is the mass ratio of the stars.

the study acknowledges its limitations of simplified numerical models and uncertainty in initial conditions for planet formation around low-mass stars. the study concludes that water delivery to habitable-zone planets around M dwarfs is unlikely due to the absence of distant giant planets.

Chapter 3

Project Vision

Imagine looking at the sky, marveling at the site of our sky and wondering could there be other worlds possibly harboring life like ours? If so, would we ever be able to go to those foreign worlds? Could there be our new possible home? That's the vision behind the Habitable Exoplanet Explorer project. To answer these broad questions, to quench the thirst of our ever growing curiosity. By combining technology with the spirit of exploration, we're opening doors too new realms of knowledge and understanding of our cosmic neighborhood. This project is not just about data and algorithms–it's about attempting to answer these questions, igniting curiosity, sparking wonder and uniting people around the world for this adventure of space exploration. Together, we're reaching for the stars and unlocking the mysteries of the cosmos.

3.1 Problem Statement

This presents the challenge of finding habitable exoplanets among the vast dataset of planets discovered by NASA's Kepler and TESS. That calls for creating sophisticated algorithms that are able to effectively process complex astronomical data. Furthermore, scientists want interfaces that are easy to use, and will help them engage with and understand the data.

3.2 Business Opportunity

We're building a powerful system to sort through this data and help scientists identify potentially habitable exoplanets. This isn't to explore our place in the universe, this is about understanding our place in the universe. Not only are we helping researchers, but by providing these insights we're sparking new collaborations and innovations in the field. It's not just a project about science, it's about creating the future of space exploration.

3.3 Objectives

1. A comprehensive system for the identification and characterization of habitable exoplanets will be developed.

2. Enhance the understanding of potential life beyond Earth through advanced data analysis and interpretation.

3. Help give valuable insight and actionable information to help future space missions and scientific endeavors.

4. Inspire public interest in space exploration and scientific discovery through accessible and engaging platforms.

3.4 Project Scope

1. Acquire data from sources such as NASA's Kepler, TESS, and relevant space missions.

2. Data preprocessing to clean, format, and prepare the collected data for analysis.

3. Analysis and visualization of exoplanet properties and characteristics to identify potential habitable candidates.

4. Development of machine learning models to predict the likelihood of exoplanet habitability.

5. Integration of the developed system into a user-friendly web application for public access and exploration.

3.5 Constraints

The main constraints of this project are:

1. Performance

2. Algorithm compatibility

3. Data Preprocessing

3.6 Stakeholders Description

Stakeholders of this project are as follow:

1. Researchers

2. Space Agencies

3. Universities, general public etc

3.6.1 Stakeholders Summary

The project involves various stakeholders, including researchers, space agencies, the public, educators, industry partners. Each stakeholder contributes unique expertise, resources, and perspectives to advance exoplanet research and exploration.

3.6.2 Key High Level Goals and Problems of Stakeholders

The primary objective is to develop a project that brings value and usefulness to the stakeholders' lives. The Project Manager assumes the responsibility of effectively managing the stakeholders, ensuring that the project aligns with its constraints while delivering tangible benefits to them. In making decisions, the project manager must exercise wisdom to prevent any negative impact on both the stakeholders and the project itself.

Chapter 4

Software Requirements Specifications

This chapter will have the functional and non functional requirements of the project.

4.1 List of Features

1. **Stellar Classification:**
   - Classify host stars based on their spectral type, luminosity, temperature, and other properties.
   - Categorize stars into different classes for further analysis.
2. **Habitable Zone Calculation:**
   - Utilize stellar properties to calculate the habitable zone for each star in the dataset.
   - Define regions where planets could potentially support liquid water based on stellar characteristics.
3. **Exoplanet Characterization:**
   - Estimate key parameters of exoplanets such as mass, radius, and orbital period.
4. **Machine Learning Model Development:**

- Train a machine learning model to predict the likelihood of habitability based on a combination of stellar and exoplanet features.

5. **Model Evaluation and Fine-Tuning:**

   - Evaluate the performance of the machine learning model using appropriate metrics.
   - Fine-tune model parameters to improve accuracy and address over-fitting or biases.

6. **Web Application:**

   - Develop an user-friendly web application to showcase habitable exoplanets and their key features.
   - Engage the public and scientific community through outreach efforts and collaborations.

4.2 Functional Requirements

ID	Requirements
FR001	Up to date data retrieved and stored in structured format
FR002	Stored data is thoroughly preprocessed
FR003	User should get correct username and password for login
FR004	User can upload their dataset
FR005	User can view predicted result
FR006	Website must have a main page after login verification
FR007	Habitability score must be shown to all the planets

Table 4.1: Functional Requirements of Habitable Exoplanet Explorer

4.3 Quality Attributes

Following are the project's quality attributes:

1. Accuracy

2. Reliability

3. Scalability

4. Usability

4.4 Non-Functional Requirements

ID	Requirements
NR001	The system must be secure from hackers
NR002	The system must responds to user queries in a specified time even under load
NR003	The system should always be available to users
NR004	If the system crashes, there must be a recovery method
NR005	Web App should be platform independent

Table 4.2: Functional Requirements of Habitable Exoplanet Explorer

4.5 Use Cases/Use Case Diagram

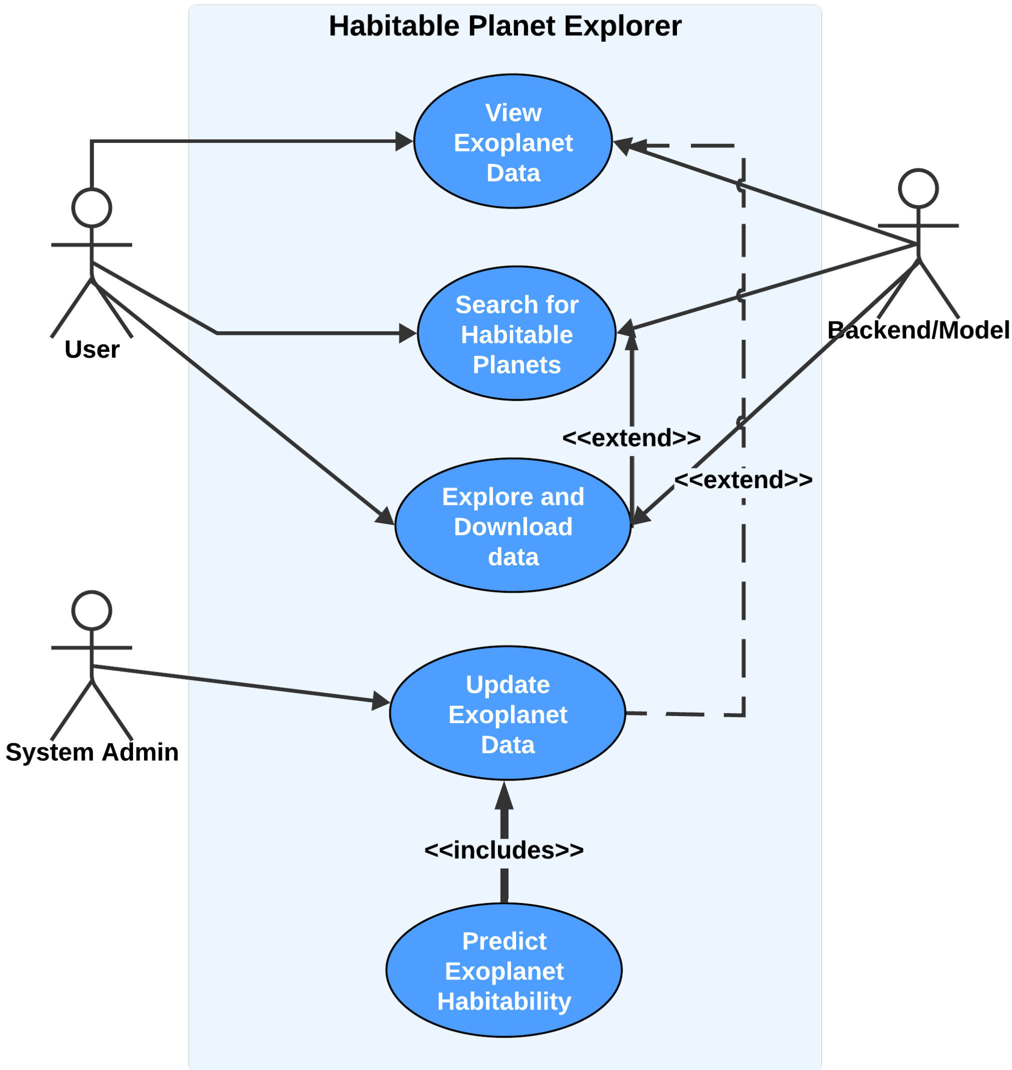

Figure 4.1: Use Case Diagram

4.6 System Diagram

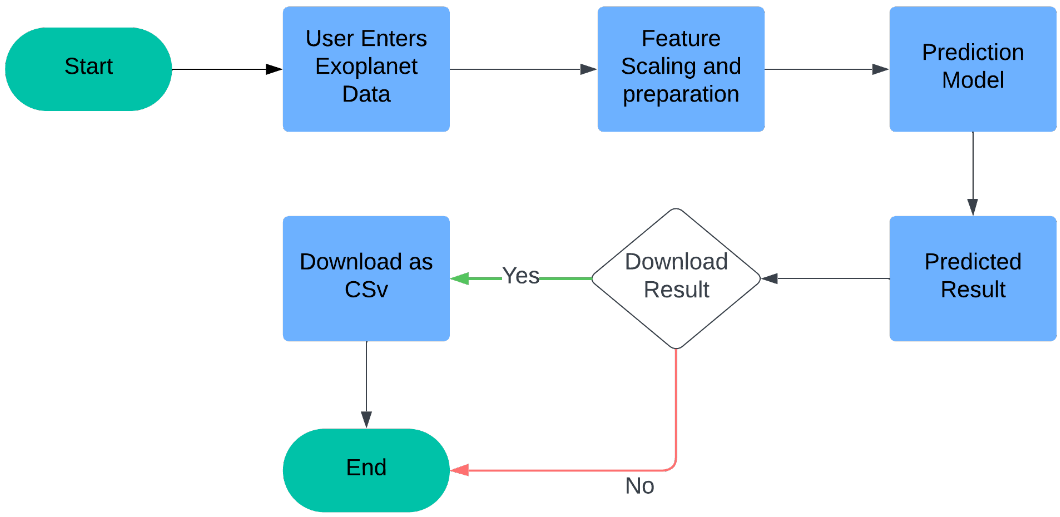

Figure 4.2: System Diagram

4.7 Test Plan (Test Level, Testing Techniques)

Test plan for our project will be as:

- Functional Testing
- Performance Testing
- Usability Testing
- Integration Testing
- Security Testing

4.8 Software Development Plan

Our software development plan will be as :

- Project Overview
- Requirements Gathering and Analysis
- System Design
- Development
- Testing
- Deployment
- Documentation
- Maintenance
- Management

Chapter 5

Iteration Plan

5.1 Planning Strategy

- FYP 1 Mid:
  - Literature Review
  - Collecting Dataset
  - Stellar Classification
- FYP 1 Final:
  - Habitable Zone Calculation
  - Exoplanet Characterization
- FYP 2 Mid:
  -
  - Model Evaluation and Fine Tuning
- FYP 2 Final:
  - Web Application Development
  - Integration and Deployment
  - Final Product

5.2 Dataset 1 (Planetary Dataset)

This is the Exoplanet dataset with flag=1 (representing only the confirmed exoplanets are selected), downloaded from NASA Exoplanet Archive, planetary systems. Below is the screenshot of the key features used for analysis.

	pl_name	hostname	pl_rade	pl_masse	pl_orbper	pl_orbsmax	st_spectype	st_teff	st_rad	pl_dens	st_met	pl_orbeccen
2	11 Com b	11 Com	NaN	NaN	323.21000	1.178	G8 III	4874.0	13.76	NaN	-0.26	0.238
3	11 UMi b	11 UMi	NaN	NaN	516.21997	1.530	NaN	4213.0	29.79	NaN	-0.02	0.080
6	14 And b	14 And	NaN	NaN	186.76000	0.775	K0 III	4888.0	11.55	NaN	-0.21	0.000
11	14 Her b	14 Her	NaN	2559.47216	1765.03890	2.774	NaN	NaN	NaN	NaN	NaN	0.373
22	16 Cyg B b	16 Cyg B	NaN	NaN	798.50000	1.660	NaN	5750.0	1.13	NaN	0.06	0.680

Figure 5.1: Planetary Systems Dataset

5.3 Dataset 2 (Direct Imaging Dataset)

This dataset explores the planets' data acquired through Direct Imaging. It's Analysis yields a different perspective on understanding the possible planets' characteristics. This dataset is downloaded from NASA Exoplanet Archive, direct imaging. Following are the key features used for its analysis:

- plntname: Planet name
- impltemp: Inferred planet temperature (K) – Directly relevant to habitability.
- implradius: Inferred planet radius
- implmass: Inferred planet mass

	plntname	ra_str	ra	dec_str	dec	imdelthmag	imdelthmagerr	imdelthmaglim	imdeltjmag	imdeltjmagerr	...	impltemperr1
0	51 Eri b	37:36.1	69.400551	-02:28:24.8	-2.473548	14.43	0.23	0.0	14.40	0.404	...	75.0
1	HIP 65426 b	24:36.1	201.150406	-51:30:16.1	-51.504459	11.14	0.05	0.0	12.70	0.400	...	100.0
2	USco CTIO 108 b	05:54.1	241.475320	-18:18:44.4	-18.312332	2.91	0.09	0.0	3.11	0.090	...	100.0
3	2MASS J01225093-2439505 b	22:50.9	20.712243	-24:39:50.7	-24.664049	6.18	0.04	0.0	5.80	NaN	...	NaN

Figure 5.2: Direct Imaging Dataset

Chapter 6

Iteration 1

This iteration comprises of in depth research, reading various papers on habitability, how planets are formed, what affects the habitability etc. Planetary Dataset is collected, basic data pre processing such as cleaning, filtering data, handling missing values etc. Further more, determining host stars' spectral type as well as calculating stellar luminosity.

6.1 Stellar Classification

st-teff: Stellar effective temperature (K): – The temperature of the star's surface, which directly affects the habitable zone location and planet's climate. Stellar Classification is classifying the stars based on their temperature. it categorizes and help us understand what class a star belongs to. for example G sequence stars, M dwarf stars etc, helps us deduce the planetary systems and get a general idea of what might be the planets' composition might be from our previous understanding.

```
# Determine Spectral Type based on st_teff
def calculate_spectral_type(temperature):
    if temperature > 30000: return 'O'
    elif temperature > 10000: return 'B'
    elif temperature > 7500: return 'A'
    elif temperature > 6000: return 'F'
    elif temperature > 5200: return 'G'
    elif temperature > 3700: return 'K'
    else: return 'M'

df_filtered['st_spectype'] = df_filtered['st_teff'].apply(calculate_spectral_type)
```

Figure 6.1: Function to determine spectral type of host star

6.2 Stellar Luminosity

Stellar luminosity refers to **the total amount of energy emitted by a star**, which plays a crucial role in determining the overall brightness and characteristics of the star. Calculated as follows:

$$L = 4\pi R^2 \sigma T_{\mathrm{eff}}^4$$

where:

- L is the luminosity,
- R is the radius of the star,
- T_{eff} is the effective temperature of the star,
- σ is the Stefan-Boltzmann constant, $5.67 \times 10^{-8}\,\mathrm{W/m^2\,K^4}$.

It help us understand how much energy do the planets receive from their host stars which influence the planets' temperature and their habitability.

Chapter 7

Iteration 2

In this Iteration, an in depth exploratory data analysis is conducted along with habitable zone calculation and advanced exoplanet characterization. Each step consists of sub-steps required to complete the iteration goals. Let's take a look at each step in depth.

7.1 Habitable Zone Calculation

Habitable zone is the distance from a star at which liquid water could exist on planets' surfaces. Where conditions might be just right – neither too hot nor too cold – for life as we know it. To calculate habitable zone for planets, we will use different approaches to give us the best estimation:

- Calculate semi-major axis using Kepler's Third Law
- Calculate Effective Stellar Flux
- Estimate equilibrium temperature with albedo consideration
- Approximate Greenhouse Factor

7.1.1 Calculate semi-major axis using Kepler's Third Law

The semi-major axis is a key parameter in determining the size of an elliptical orbit of a planet. Using this, we can determine the average distance between a planet and its host star. According to Kepler's Third Law, the square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit. This law helps us understand how far a planet is from its star and how long it takes to complete one orbit. In the context of habitability, the semi-major axis can help determine the habitable zone (HZ)

```
def calculate_semi_major_axis(period, star_mass):
    # Kepler's Third Law: a^3 = (P^2 * G * (M + m)) / (4 * pi^2)
    # Assuming planet mass (m) is negligible compared to star mass (M)
    G = 6.67430e-11  # Gravitational constant in m^3 kg^-1 s^-2
    period_seconds = period * 86400  # Convert days to seconds
    star_mass_kg = star_mass * 1.989e30  # Convert solar masses to kg
    semi_major_axis_meters = (G * star_mass_kg * period_seconds**2 / (4 * np.pi**2))**(1/3)
    semi_major_axis_au = semi_major_axis_meters / 1.496e11  # Convert meters to AU
    return semi_major_axis_au
```

Figure 7.1: Function to calculate semi major axis of a planet.

7.1.2 Calculate Effective Stellar Flux

The Effective Stellar Flux is the amount of energy per unit area received by a planet from its host star. It helps us determining the planet's surface temperature which plays a critical role in a planet's habitability. The effective stellar flux depends on the star's luminosity and the distance from the star, which we have already calculated.

```
# Calculate Effective Stellar Flux (S_eff) and ensure it's positive
df_filtered['s_eff'] = (df_filtered['luminosity'] / (4 * np.pi * df_filtered['pl_orbsmax']**2)).abs()
```

Figure 7.2: Function to calculate stellar flux of a planet.

7.1.3 Estimate equilibrium temperature with albedo consideration

The Equilibrium Temperature of a planet is the temperature at which the energy received from its host star (through stellar flux) is balanced by the energy the planet radiates back into space. This equilibrium balance becomes the the ultimate decider whether the planet's temperature is compatible in hosting liquid water. The planet's albedo is the decider of the amount of energy which is reflected by a planet. A higher albedo means the planet reflects more energy, resulting in lower temperature.

```
def calculate_equilibrium_temp_with_albedo(s_eff, albedo=0.3):
    sigma = 5.67e-8  # Stefan-Boltzmann constant
    s_eff_abs = abs(s_eff) # Take absolute value to avoid negative fourth roots
    return ((1 - albedo) * s_eff_abs / (4 * sigma))**(1/4)
```

Figure 7.3: Function to calculate equilibrium temperature of a planet.

7.1.4 Approximate Greenhouse Factor

Greenhouse factor can be used to further determine the planet's temperature. Since the dataset doesn't contain atmospheric data, so we will use a simplified approach using planet's mass and density. Typically larger planets with lower density (such as gas giants or mini-Neptunes) are likely to have a stronger greenhouse effect. On the other hand, smaller planets (such as Earth-like planets) are assumed to have a weaker or minimal greenhouse effect.

$$G = \begin{cases} 1.5 & \text{if } r > 1.5\,R_\oplus \text{ and } d < 2\,\text{g/cm}^3 \quad \text{(Gas Giant, strong greenhouse effect)} \\ 1.2 & \text{if } r > 1\,R_\oplus \text{ and } d < 3\,\text{g/cm}^3 \quad \text{(Mini-Neptune, moderate greenhouse effect)} \\ 1.0 & \text{if } r \leq 1\,R_\oplus \text{ and } d \geq 3\,\text{g/cm}^3 \quad \text{(Earth-like, minimal greenhouse effect)} \end{cases}$$

The habitable zone boundaries are then adjusted based on the greenhouse factor as follows:

$$\text{hz\_inner} = \text{hz\_inner} \times G^{-0.5}$$

$$\text{hz\_outer} = \text{hz\_outer} \times G^{-0.5}$$

7.2 Exoplanet Characterization

Let's categorize the planets based on key properties. Mass and radius with the following ranges:

For planet's Radius characterization:

$$\text{Radius Category} = \begin{cases} \text{Earth-sized} & \text{if } r < 1.25 R_{\oplus} \\ \text{Super-Earth} & \text{if } 1.25 \leq r < 2.0 R_{\oplus} \\ \text{Neptune-like} & \text{if } 2.0 \leq r < 6.0 R_{\oplus} \\ \text{Gas Giant} & \text{if } r \geq 6.0 R_{\oplus} \end{cases}$$

For planet's Mass characterization:

$$\text{Mass Category} = \begin{cases} \text{Low-mass} & \text{if } m < 2.0 M_{\oplus} \\ \text{Medium-mass} & \text{if } 2.0 \leq m < 10.0 M_{\oplus} \\ \text{High-mass} & \text{if } m \geq 10.0 M_{\oplus} \end{cases}$$

Characterization Output:

```
Size Category Counts:
| size_category   | count   |
|:----------------|:--------|
| Neptune-like    | 3244    |
| Super-Earth     | 1070    |
| Gas Giant       | 813     |
| Earth-sized     | 511     |

Mass Category Counts:
| mass_category   | count   |
|:----------------|:--------|
| High-mass       | 5294    |
| Medium-mass     | 298     |
| Low-mass        | 46      |
```

Figure 7.4: Planet's distribution based on their category

Chapter 8

Iteration 3

In this iteration, a deep learning model will be developed to predict the habitability of a given planet with confidence. The activation function chosen for the given task is **SELU** as it yielded the best results because it enhances convergence speed, especially for our dataset with complex patterns.

8.1 Data Preparation for the model

8.1.1 Splitting Dataset

Dataset has been split into training and testing sets with the following dimensions:

```
#########Summary###########
X_train shape: (3031, 41)
X_test shape: (758, 41)
Y_train shape: (3031,)
Y_test shape: (758,)
###########################
```

Figure 8.1: Split data dimensions

8.1.2 Feature Scaling

Dataset has been normalized the dataset to be fed to the model. Furthermore, scaler used to normalized the dataset has been saved to be used in real-time prediction in the Web App.

8.2 Model Building

8.2.1 Model definition

After trial and test, following structure of the network has been selected.

```
classifier=Sequential()
classifier.add(Dense(units=20,kernel_initializer='he_uniform',activation='selu',input_dim=41))
classifier.add(Dropout(0.5))
classifier.add(Dense(units=12,kernel_initializer='he_uniform',activation='selu'))
classifier.add(Dropout(0.3))
classifier.add(Dense(units=12,kernel_initializer='he_uniform',activation='selu'))
classifier.add(Dropout(0.4))
classifier.add(Dense(units=3,kernel_initializer='glorot_uniform',activation='softmax'))
classifier.compile(optimizer = 'Adamax', loss = 'sparse_categorical_crossentropy', metrics = ['accuracy'])
```

Figure 8.2: Neural Network Structure

Neural network consisting of three hidden layers using **SELU** activation, and **He uniform** initialization to get uniform initial weights, which would optimize weight scaling and avoid exploding, vanishing gradients. Dropout layers(0.3-0.5) to reduce any over-fitting.

8.2.2 Model Accuracy

Following image shows the accuracy of the model during its learning phase.

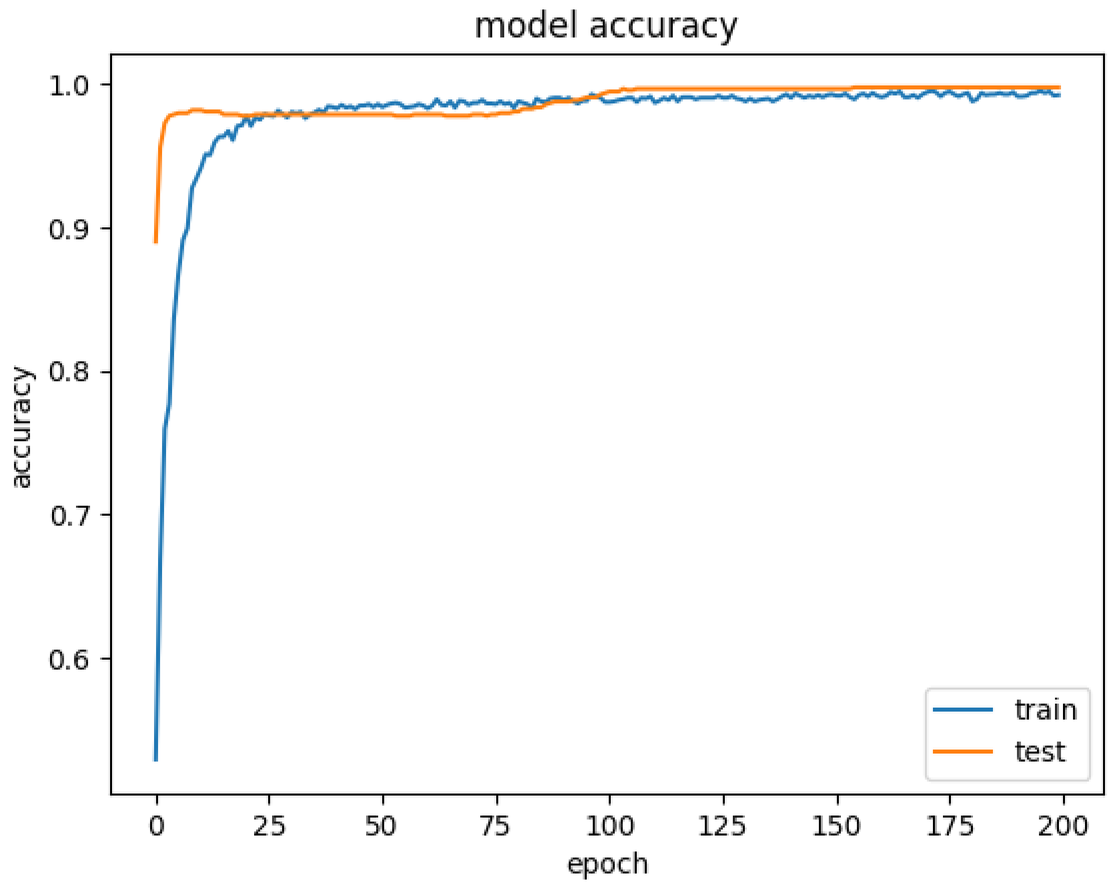

Figure 8.3: Model Accuracy during learning phase

8.2.3 Model Loss

Following image shows the loss of the model during its learning phase.

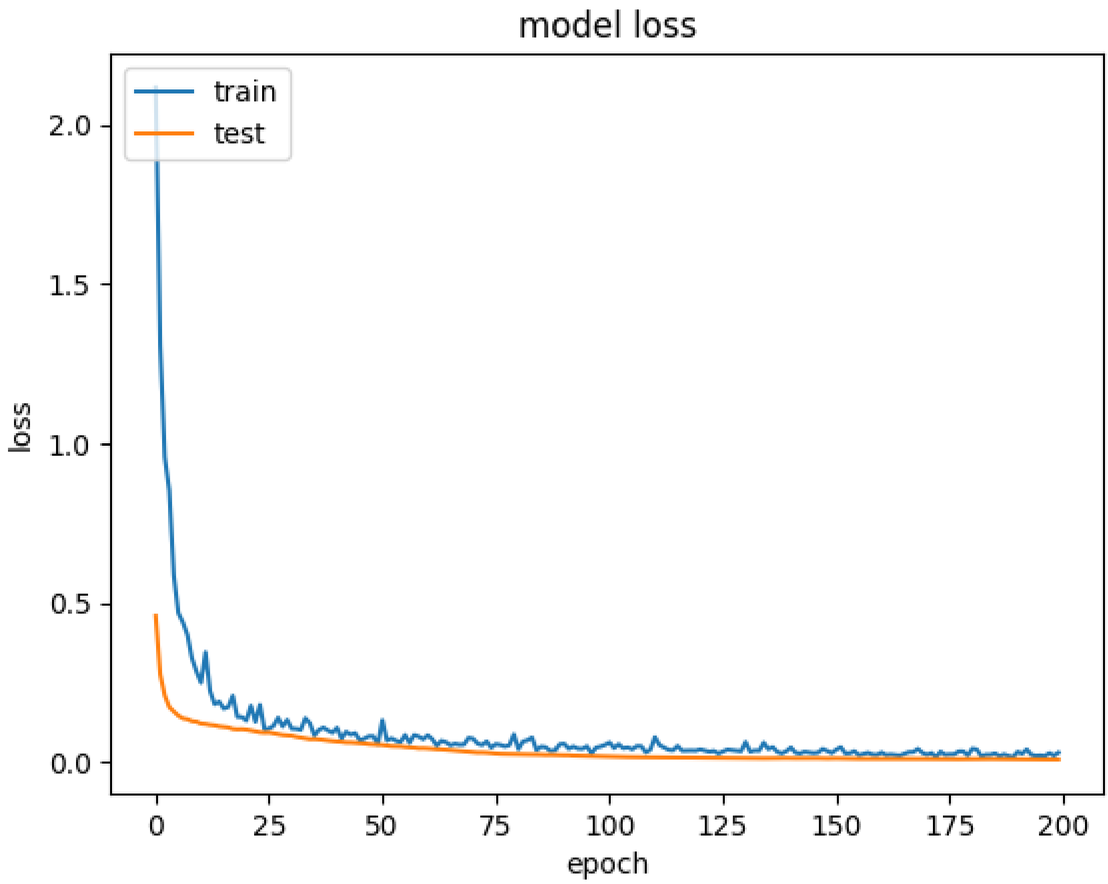

Figure 8.4: Model Loss during learning phase

8.2.4 confusion matrix

Following image shows the confusion matrix of the predictions made by the model

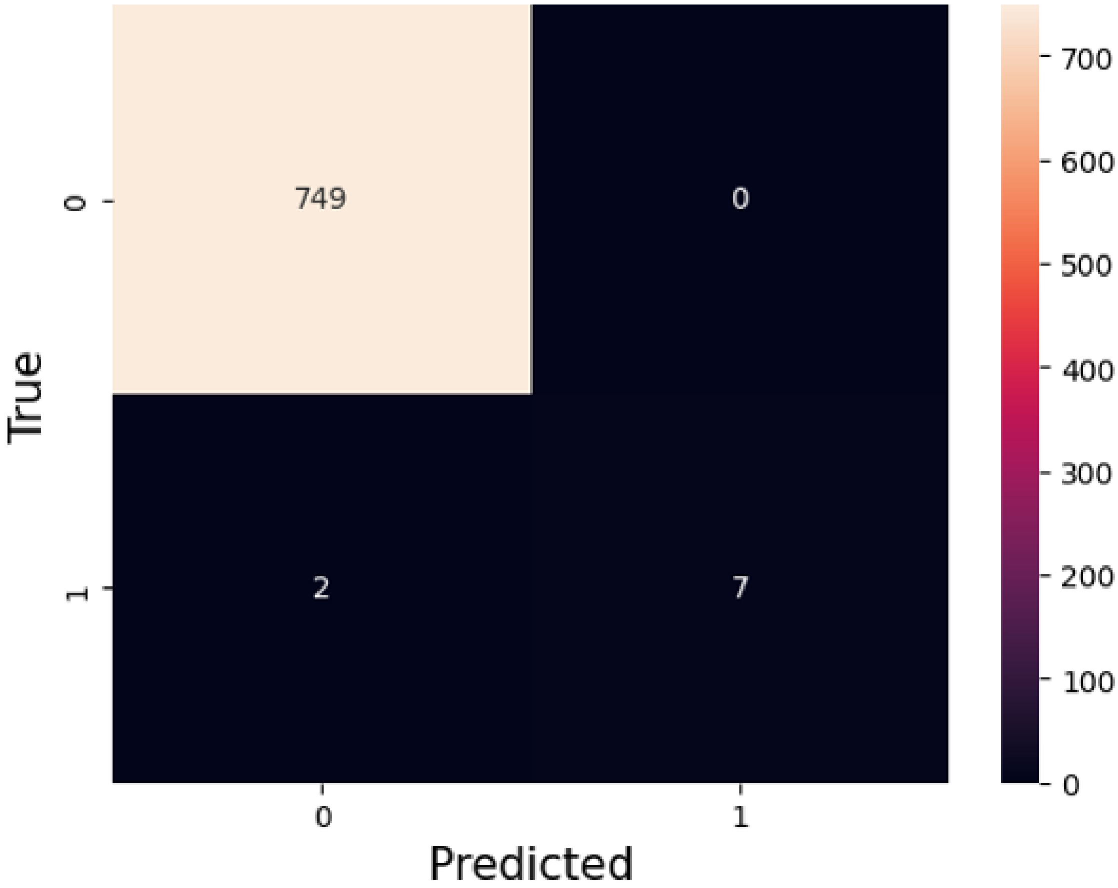

Figure 8.5: confusion matrix

8.2.5 Model metrics

Following are the model metrics. The model predicts exceptionally well on the habitability of the planet.

F1 Score: [0.998 0.975] Recall: [1. 0.969] Precision: [0.997 1.] Accuracy: 0.997

Chapter 9

Iteration 4 (Final Web Application

9.1 User Manual

9.1.1 Homepage

The development of the web application was successfully completed. The web application has been integrated with the back end working model, providing a functional system.

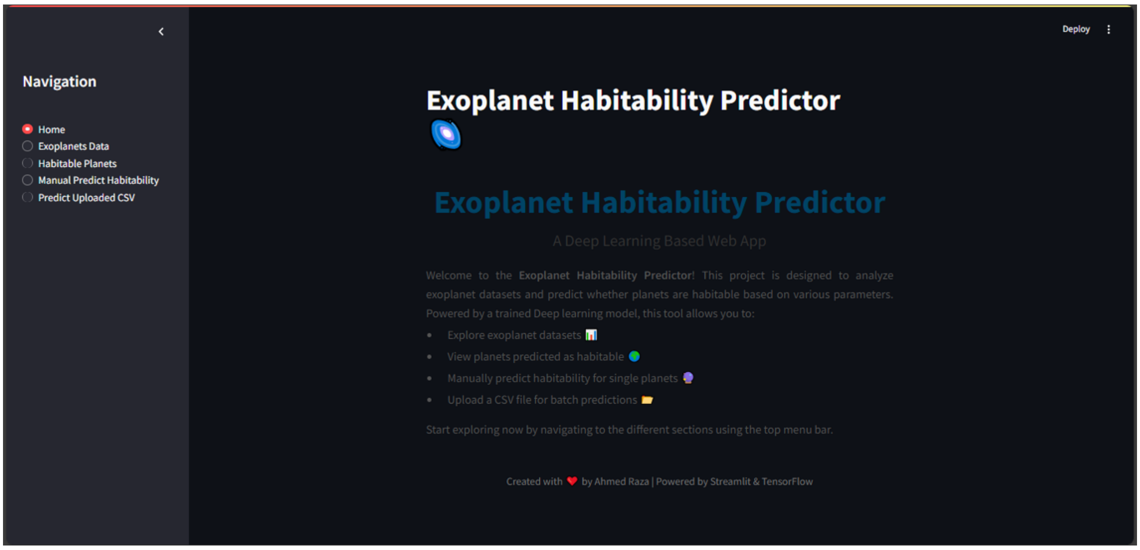

Figure 9.1: Web Application (Home page)

9.1.2 Exoplanet page

Here, the dataset is completed and ready to be viewed and downloaded in csv file.

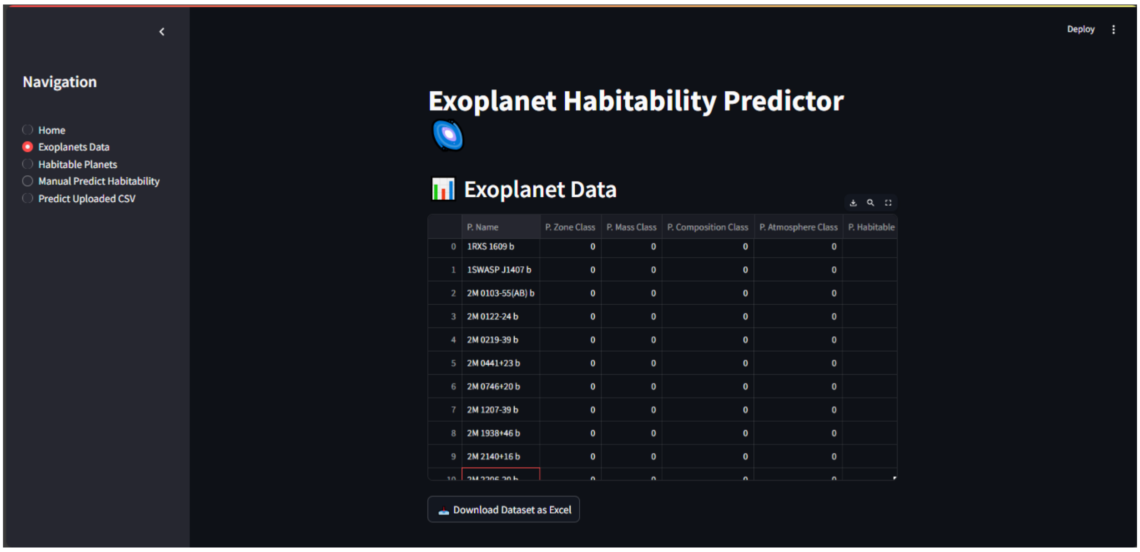

Figure 9.2: Web Application (Explore Exoplanet page)

9.1.3 Exoplanet page

Here, the dataset consists of the planets, our model predicted to be habitable with confidence. It can be downloaded

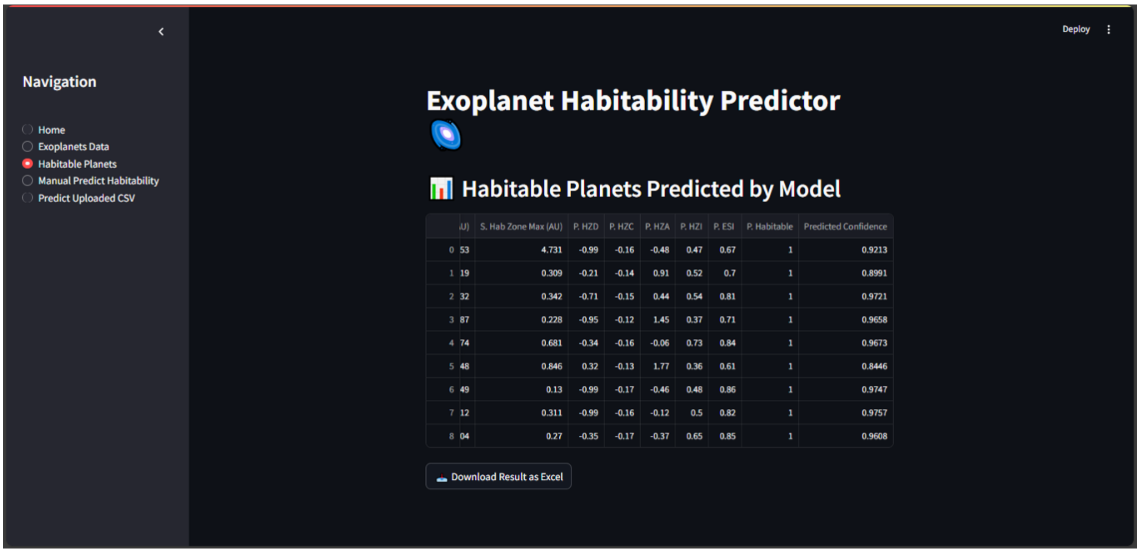

Figure 9.3: Web Application (Habitable Exoplanet page)

9.1.4 Manual Predict Habitability

On this page, a sandbox like experience is embedded, where you can input various values and see if any of those affect the habitability on your inputted planet.

Figure 9.4: Web Application (Manual Predict Habitability 1)

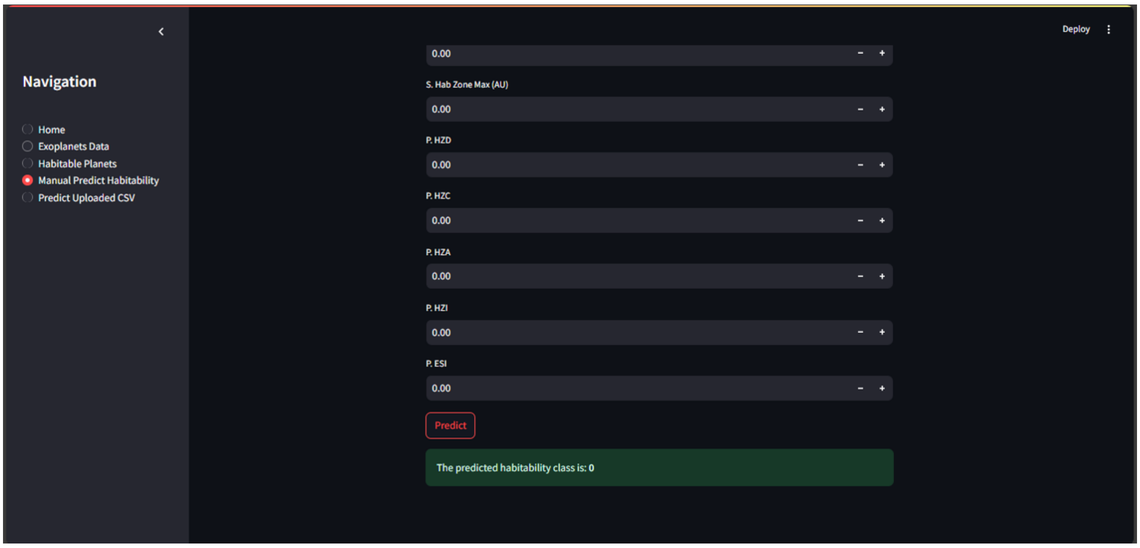

Figure 9.5: Web Application (Manual Predict Habitability 2)

9.1.5 Predict Uploaded CSV

Here, You can upload the csv file with multiple planets to the application and the model will assess each of them and provide the output with prediction and confidence. which you can then download as csv file.

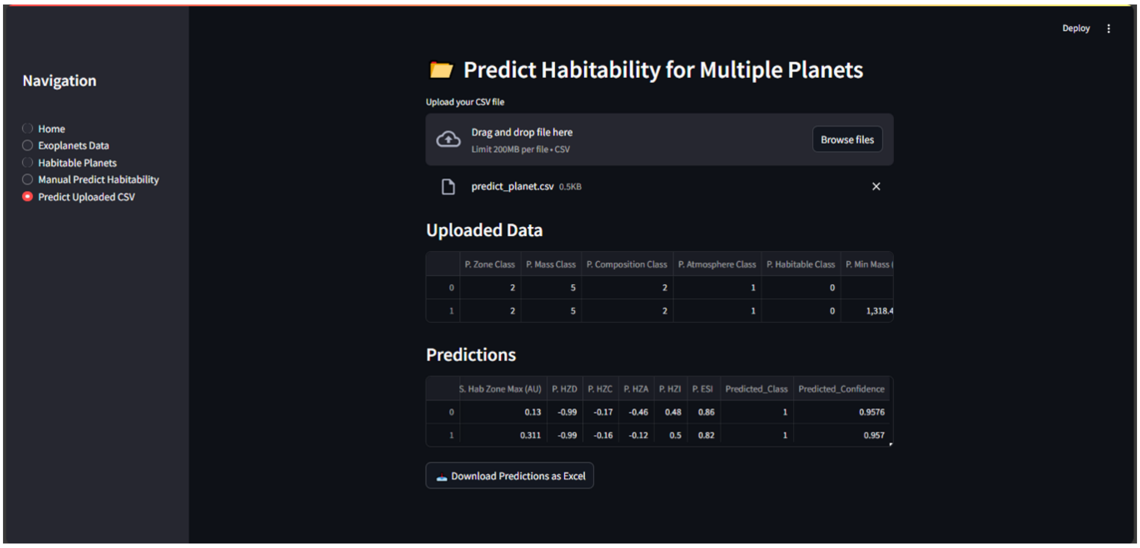

Figure 9.6: Web Application (Predict Uploaded CSV page)

9.2 Link to the Application

Following is the link to the application:

Habitable Exoplanet Explorer: https://habitableexoplanet-h6gbnnlv2w4fwr6s4gwvqn.streamlit.app/

Chapter 10

Conclusions and Future Work

10.1 Conclusion

This project successfully integrated multiple advanced approaches and methodologies to explore the habitability of exoplanets. We derived various planetary and stellar properties such as equilibrium temperature, stellar flux, stellar luminosity, greenhouse factor, exoplanet characterization etc, which proved to be good indicators for habitability. The deep learning model SELU-based proved exceptionally well in delivering highly accurate predictions of potentially habitable exoplanets.

10.2 Future Work

For future research, we could benefit from integrating spectroscopic and atmospheric data of exoplanets to further enhance our understanding about the habitability of exoplanets, when available. Current thresholds (e.g., mass, radius, temperature) could be expanded to include more advanced factors like atmospheric composition and magnetic field strength.

Bibliography

Steve Bryson, Michelle Kunimoto, Ravi K Kopparapu, Jeffrey L Coughlin, William J Borucki, David Koch, Victor Silva Aguirre, Christopher Allen, Geert Barentsen, Natalie M Batalha, et al. The occurrence of rocky habitable-zone planets around solar-like stars from kepler data. *The Astronomical Journal*, 161(1):36, 2020.

Matthew S Clement, Elisa V Quintana, and Billy L Quarles. Habitable planet formation around low-mass stars: rapid accretion, rapid debris removal, and the essential contribution of external giants. *The Astrophysical Journal*, 928(1):91, 2022.

Rudolf Dvorak, Thomas I Maindl, Christoph Burger, Christoph Schäfer, and Roland Speith. Planetary systems and the formation of habitable planets. *arXiv preprint arXiv:1502.02937*, 2015.

Brendan Dyck, Jon Wade, and Richard Palin. The effect of core formation on surface composition and planetary habitability. *The Astrophysical Journal Letters*, 913(1):L10, 2021.

Nader Haghighipour and Sean N Raymond. Habitable planet formation in binary planetary systems. *The Astrophysical Journal*, 666(1):436, 2007.

Franck Selsis, James F Kasting, Benjamin Levrard, Jimmy Paillet, Ignasi Ribas, and Xavier Delfosse. Habitable planets around the star gliese 581? *Astronomy & Astrophysics*, 476(3):1373–1387, 2007.

Philippe Thebault, Francesco Marzari, and Hans Scholl. Planet formation in the habitable zone of α centauri b. *Monthly Notices of the Royal Astronomical Society: Letters*, 393 (1):L21–L25, 2009.