THE ANNUAL TECHNO-MANAGEMENT FEST, IIT (BHU) VARANASI

TECHNEX'25

28 FEB - 2 MAR 2025

STELLAR ANALYTICS

INTRODUCTION

Prelude: The Fall of Earth

The year is **2224**, but the seeds of Earth's downfall were sown centuries earlier. In the 21st and 22nd centuries, humanity thrived, reaching unprecedented technological and economic heights. However, these achievements came at a devastating cost:

- **2050s**: The first cracks appeared as global temperatures soared, ice caps melted, and extreme weather events ravaged cities.
- **2100s**: Oceans rose to swallow coastal regions, agricultural systems collapsed, and freshwater resources dwindled. Despite warnings, industrial practices continued, pumping carbon dioxide into an already fragile atmosphere.
- **2150s**: The final blow came with the depletion of critical resources. Desperate geoengineering experiments to reverse climate change spiraled out of control, destabilizing the atmosphere and rendering vast regions uninhabitable.

By **2200**, Earth was no longer the vibrant cradle of life it once was. Starvation, wars over dwindling resources, and mass migrations left humanity fractured and desperate. Scientists and leaders agreed: **Earth was beyond saving.**

The Stellar Project: Humanity's Last Hope

In the face of annihilation, the world's remaining nations formed an unprecedented alliance — **The Stellar Coalition**. Together, they launched The Stellar Project, a monumental effort to secure humanity's survival by colonizing habitable planets beyond the solar system.
Over decades, advanced probes and telescopes scanned the cosmos, identifying hundreds of exoplanets. Yet, not all these planets are capable of supporting life. The Stellar Coalition has tasked you, the brightest minds of the next generation, with the critical mission:

Identify which exoplanets can sustain life and serve as humanity's new home.

THE ANNUAL TECHNO-MANAGEMENT FEST, IIT (BHU) VARANASI
TECHNEX'25
28 FEB - 2 MAR 2025

GENERAL INSTRUCTIONS

1. **Dataset Access:**
   - The dataset for the problem can be found here: exoplanet_dataset
   - To know about the column headings, refer to: exoplanet_dataset_info.

2. **Main Source of Dataset:**
   - The main source of the dataset for this problem is the Exoplanet Archive. (You may also refer to the documentation from this site.)

3. **Additional Resources:**
   - To know more about habitability:
     - Planetary Habitability - Wikipedia
     - Planetary Habitability
     - How Scientists Search for Habitable Planets
     - A More Comprehensive Habitable Zone for Finding Life on Other Planets
   - To know more about the Earth Similarity Index (ESI):
     - Earth Similarity Index

4. **Problem Breakdown:**
   - The problem is divided into three parts:
     - Data Preprocessing and Exploratory Data Analysis (EDA)
     - Feature Engineering
     - Creating a Custom "**Habitability Index**"

5. **Programming Environment:**
   - Participants are free to use any programming language, environment, and library. However, it is preferable to use Google Colab or Jupyter Notebook as the programming environment.

OVERVIEW

The task involves cleaning and preprocessing the raw data, deriving meaningful features such as the ESI and others (which can be derived from the dataset) that might be useful for the custom Habitability Index, addressing missing or noisy values with well-justified methods, and visualizing the relationships between features.

Participants will then design a custom "Habitability Index" using features (directly from the dataset or derived from dataset fields) like ESI, long-term stability, and atmospheric retention to rank planets based on their suitability for sustaining life. Building on this analysis, participants will develop a model to classify planets into three categories: potentially habitable (Earth-like), extreme but marginally habitable (requiring adaptation), and non-habitable (inhospitable environments).

Round 2 will culminate in the development of an interactive dashboard to present findings, including habitability trends, classification results, and visual insights. Emphasis will be placed on scientific reasoning, model accuracy, and the clarity of insights presented.

PROBLEM STATEMENT

ROUND 1

TASKS

A. Data Preprocessing & Analysis of Dataset:

1. Analyze the different fields in the dataset and identify redundant fields. Provide reasons for their elimination.
2. Address outliers and null values
3. Identify and rectify false positives, such as non-existent planets or any other inconsistent data entries.

B. Feature Engineering:

1. Feature Reduction:
   - Visualize the correlation matrix to identify redundant or highly correlated features within a dataset containing numerous features.
   - Select a suitable feature reduction technique for this dataset and elaborate on the rationale behind the chosen method.
2. Derive or estimate features such as the Earth Similarity Index (ESI), long-term stability, and atmospheric retention for use in your custom "Habitability Index."
   - Ensure that the scientific reasoning behind these engineered features is well-documented, emphasizing their importance to the Habitability Index.

C. Exoplanet "Habitability Index":

1. Build a multi-class classifier to categorize planets into three categories:
   - **Potentially Habitable**: Conditions conducive to liquid water.
   - **Marginally Habitable**: Extreme but not impossible conditions for life.
   - **Non-Habitable**: No realistic possibility of life as we know it.
2. You can either build a Machine Learning model or adopt a Statistical model approach to create your "Habitability Index" using features derived from the datasets or directly from the datasets.
3. It is recommended to use Google **Colab** or **Jupyter Notebook** to develop your model (other notebooks are also allowed).
4. Create visualizations to showcase the classification results, habitability trends, and visual comparisons of planets in your code notebook.

SUBMISSION FORMAT

PDF DOCUMENT:

The PDF should include:

- **Title Page**: Project title, name/team, date, and contact information.
- **Problem Statement**: Brief summary of the problem and objectives.
- **Assumptions**: Key assumptions made during the solution process.
- **Custom Habitability Index**:
  - Definition/formula of the index.
  - Explanation of feature relevance and weights.
- **Approach**:
  - Preprocessing and feature engineering steps.
  - Modeling methodology to develop the habitability index.
- **Findings**: Summary of results and recommendations.
- **Appendix**: Optional charts, references, or datasets.

Note:

- **Ensure to share the link to the Jupyter/Colab notebook with viewing permissions enabled for everyone in the PDF document too.**
- **Submissions without the link will not be considered for presentation.**

ROUND 2:

An Effective Dashboard for Stellar Analytics 2025

An effective dashboard for Stellar Analytics 2025 should be visually engaging, scientifically informative, and user-friendly. It should clearly communicate the analysis, results, and insights while allowing viewers to explore the data interactively.

Key Elements to Include:

1. **Introduction and Context:**
   - **Mission Brief**: A brief overview of the goal.
   - **Dataset Summary**: Key statistics about the dataset, such as the number of planets analyzed, key features (e.g., mass, temperature), and notable gaps or challenges encountered during preprocessing.

2. **Key Visualizations and Insights:**
   - **Planetary Features Distribution**: Visualizations like histograms or box plots for features such as planetary radius, mass, and surface temperature; heatmaps showing correlations between critical parameters (e.g., stellar luminosity vs. planetary distance).
   - **Visualizations of Derived Features**: Using scatter plots or heatmaps to highlight promising planets as per the Earth Similarity Index (ESI), etc.
   - **Custom Habitability Index Breakdown**: Displaying a ranked list or bar chart of planets based on their custom "Habitability Index" scores, etc.

3. **Classification Results:**
   - **Categorization Summary**: Displaying the classification of planets into the three categories: Potentially habitable (Earth-like), Marginally habitable (challenging environments), Non-habitable (hostile environments).
   - **Model Performance Metrics**: Detailing the performance of the classification model.

Bonus Tasks:

- **Interactive Table or Cards**: Summarizing key details for the most habitable planets.
- **Hover-Over Insights**: Enabling interactive hover-over features to display details of a specific planet when clicked.

Note:

There will be an element of surprise in Round 2. Participants who have qualified in Round 1 may be required to incorporate additional fields into their provided dataset. Subsequently, they will need to identify which parameters of their solution will be impacted and make the necessary adjustments.