STELLAR ANALYTICS

Description:

Welcome to Stellar Analytics, a captivating event within Technex! The Astronomy Data Analysis Event is a hands-on competition designed to challenge participants in the field of astronomy by providing them with real astronomical data sets to analyze and draw scientific conclusions from.

Structure:

The event consists of two rounds:

Round 1 : Participants are provided with raw astronomical data, often obtained from telescopes, space missions, or observatories. The task is to retrieve, clean, and preprocess the data to make it suitable for analysis. In this round, participants formulate hypotheses related to the data and perform statistical tests or data analysis techniques to test these hypotheses. Submissions will be evaluated.

Round 2 : The teams selected from the first round will be invited to an onsite presentation. Participants must communicate their results effectively to judges and potentially a broader audience.

Note : The number of teams advancing to the second round will be determined based on participation and performance in the first round.

Eligibility:

1. All students from authorized institutions are eligible to participate.
2. The Team Size limit is a maximum of 3. The participants in a team may be from the same or different institutions and programs.
3. Participation is restricted to college and university students only.
4. Bring a physical ID card, which is compulsory for all college students.

Introduction

In the distant reaches of the cosmos, a Type 2 civilization thrived on the brink of technological mastery. Their insatiable hunger for energy led them to a groundbreaking discovery: the ability to extract immense power from the cataclysmic heat of supernovas.

Driven by this revelation, the civilization launched a grand endeavor to identify the supernovas most conducive to energy extraction. Their astronomers and scientists delved deep into the heart of these cosmic explosions, deciphering the complex interplay of supernova types, luminosity, and spectral characteristics to pinpoint the ideal candidates for their revolutionary technology.

Yet, this quest was more than a mere pursuit of energy; it was a profound journey into the cosmos, where the civilization sought to harmonize with the natural forces of the universe. Their vision extended beyond technological prowess, envisioning a future where their society thrived on the boundless energy of the cosmos, forging a new era of sustainability and advancement.

As their understanding of supernovas deepened, so did their reverence for the cosmic ballet that governed the universe. Each discovery brought them closer to a future where they would stand as stewards of cosmic energy, harnessing the raw power of the universe itself to propel their civilization towards a harmonious coexistence with the cosmos.

General Instructions

1.The dataset for the problem can be found here : astrotrain.xlsx.

2.To know more about the types of civilizations :
https://futurism.com/the-kardashev-scale-type-i-ii-iii-iv-v-civilization

3.To know more about the different types of supernovas :
https://www.astronomy.com/science/the-different-types-of-supernovae-explained/

4.Some field descriptions can be found here :
https://zenodo.org/records/3265189
https://heasarc.gsfc.nasa.gov/docs/tess/LightCurve-object-Tutorial.html

5.The problem set contains 3 questions, each having subparts.

6.Participants are free to use any programming language, environment and library.
However, it is preferable to use Google Colab as the programming environment.

7.The submitted code will be evaluated on a predetermined validation dataset.

PROBLEM SET

I. Visualization and Analysis of the dataset:

1. Inspect the data types and extract basic statistics:
   a. Print Range, Mean, Median, and Standard Deviation of the dataset.
   b. Does the dataset require normalization?
   c. What are your inferences about the above results?
2. Using the seaborn module, plot a heatmap to explore the relation between various features of the dataset.
   a. What do you infer from the heatmap?
3. Determine the interquartile range and skewness of the dataset.
4. Report the value of the total flux by taking the mean squared values of the different flux values.
5. Calculate the cartesian coordinates using the values of right ascension and declination angles. Add new columns to the dataset for these values.

II. Feature engineering:

1. Determine the percentage of missing values for each feature and use visualization to identify features with the highest null percentages.
2. Feature reduction:
   a. Visualize the correlation matrix to pinpoint redundant or highly correlated features within a dataset containing numerous features.
   b. Select a suitable feature reduction technique for this dataset and elaborate on the rationale behind the chosen method.
   c. Examine the relationships between various features through scatter plots before and after implementing feature reduction. Observe changes in data distribution, patterns, and clusters to assess the impact of the reduction process.
3. Utilize an appropriate imputation method to address the null values in the dataset. Provide a rationale for the chosen imputation technique.

III. Supernovae Classification:

1. Build a robust and efficient classifier to classify supernovas into 7 different types:
   - Ia type
   - IIP type
   - IIn type
   - IIL1 type
   - IIL2 type
   - Ib type
   - Ic type
2. Plot the ROC curve and Confusion Matrix to quantify the performance of your classifier.
3. Fine-tune all the hyperparameters in the classifier by selecting a suitable optimization method. Justify the choice of the optimization method based on its underlying rationale.
4. Report the effect of different cross validation techniques.

INSTRUCTION FOR SUBMISSION:

1. The First Round deadline for submission is 10th March 2024. (**tentative**)
2. Participants must register on the Technex Website, and the solutions should be submitted in the format mentioned below through the unstop portal: Here
3. The solution should be submitted in PDF format only. The file's name should be SA#1_<TEAM_NAME>_SOLUTION. The language should be ENGLISH only.
4. The decision of the judges shall be final and binding.
5. The Timings for Round 2 will be from 15th March 2024 - 17th March, 2024. (**tentative)**
6. The submission format should be as follows:
   - The team's name, details of team members (name, contact number), and college name should be mentioned on the first page.
   - Solutions should be brief and to the point. The challenges encountered should be mentioned. Any figures, diagrams (well-labeled), and charts mentioned should be properly referenced in the text.
   - If you feel the figures, diagrams, and/or charts are losing detail while scaling down in PDF, make a ZIP or RAR file with the solution. The file should be named the same as the PDF and submitted to info.astronomyclub@gmail.com. Highlight the innovation in your solution.
   - The solution should not exceed ten pages.
7. Problem Statement is subject to changes.

***Terms and Conditions Apply**

EVENT COORIDNATORS:

Sonam Patel : +91 6387283487
Vansh Gupta : +91 6239084133

*Event organisers have the authority to do amendments in rules and PS