Inter IIT PrepAThon: Astro + SDE

Introduction

In InterIIT it's never about only SDE or only Astronomy. It's always about how complete of a product you are presenting. This year we are launching a SDE + Astro problem statement for the students to explore. Carefully read the whole problem statement to avoid misunderstanding. All the evaluation criteria are also mentioned in detail for your reference. Kindly note the following

- **Only Astro - People who have chosen only Astro track should complete only PSA and try to work on PSB for grace points.**
- **Only SDE - People/teams who have chosen only SDE track should complete only PSB and try to work on PSA for grace points. You should work on the Pure-SDE implementation track mentioned in PSB**
- **Astro + SDE - People/teams who have chosen Astro + SDE should complete PSA and PSB both and integrate them together as described.. The Pure SDE Implementation part should be skipped here.**

Problem Statement

PSA [Astro Track]

Introduction:
Cosmic sources in the sky, including our star, the Sun, burst intermittently in the X-ray energy band of the electromagnetic spectrum. The burst amplitude and the duration vary depending on the source and the cause of the burst. The observed burst has a fast rise and slow decay pattern/shape with variable rise and decay times. An automatic identification system for such data will further simplify the analysis process and the associated scientific investigation of these bursts.

Build a process to identify and categorize X-ray bursts. Given data will include long-duration X-ray light curves as well. Parameters like the rise and decay time, peak flux, duration of bursts, etc., must be derived. False detection should be minimum, and all bursts must be detected.

The problem can be broadly categorized into two parts:

1. Developing a statistical/machine learning model to cover the mentioned parameters.
2. Deploying it efficiently to a stand-alone application and web-based tool (no additional APIs used) (refer to PSB section for the flow details)(only for Astro + SDE people)

You should visualize the light curve data, feature a GUI for light curve distribution visualization, analyze wavelets, display particular portions of distributions as time series: confirming burst occurrence, etc.

Basic requirements:

- Light curves of X-ray source(s) are attached in the Resources section, with many outbursts and systematic variations in the background. The tool should be designed to accept any X-ray light curve with a defined format. It should allow different input file formats – like ASCII, FITS, and XLS.
- The model should identify the bursts and their properties (as listed in the problem statement) along with the time of occurrence. Classification of the bursts is to be done, and the use of fit parameters should be justified. Classification criteria is open for innovation.
- The tool should be developed completely using open source (Python/Perl etc.) software.
- Additionally, the teams must use the available data and prove the efficiency of the system w.r.t. minimal false alarms and maximal true signatures.
- The teams must bring out the limitations of the method(s) used in the tool and possible enhancement later through research.

Software specifications:

- Should be made as a script for Pure Astro people or as a web based application for Astro + SDE people. (Refer to PSB for more details about Astro + SDE Track)
- The tool should be distributed under GPL or any other open source license.
- Should use only open source, Python is desirable.

Evaluation:

For statistical/machine learning model:

- Detection of the maximum number of bursts with fewer false detections for a given duration [20 Points]
- Identification of the fit parameters (like the rise time, peak flux, decay time constant) and its accuracies: good accuracy and goodness of fit for many bursts [30 Points]
- Classification of the fit parameters and their explanations [20 Points]

Adaptability to take any format input data (FITS, ASCII, CDF, etc.) [20 Points]
Concise and clear documentation, including inferences derived for the sample data, should be available along with the codes for user demonstration [10 Points]

[Total 100 Points]

Note: Physical explanation for the use of fit parameters and the classification criteria will be prioritized for final evaluation. Also, some more resources are provided in the Resources section to help you get started!

PSB [SDE Track]
Develop a portal which can be used to perform the following things

- Users can login and signup (using Google and any one other social media). This should also include 2FA, password updation and recovery [5 (Google login) + 5 (other social media login) + 10 (2FA) + 5 (password updation) + 5 (password recovery) = 30 Points]
- Users have a UI to run computations as described in **PSA** and view results of the process by plotting interactive graphs and tables. Note that if you don't have the knowledge about Astro then follow the next section on Pure SDE Implementation. [20 Points]
- A UI to view all the previous submitted datasets/searches (searches for pure SDE track) (Basically history feature) [10 Points]
- A scalable architecture to handle DAU of minimum 1,00,000 resulting in at least 1,000 requests per second. [15 (Backend + Frontend Architecture including deployment strategies) + 25 (Backend + Frontend Implementation) = 40 Points]

TOTAL : 100 POINTS

Pure SDE Implementation: [Astro peeps should skip this section]
Download the data provided in the resources section.

1. Take input the name/code of the company from the frontend and follow the steps:Fetch the companies which partially/fully match the user input
2. Show the results on the frontend and let the user select the company
3. Perform the following computation **in real time**: (Precomputation not allowed for the first time the user clicks it. After that you serve a precomputed result when the user opens this result from the history/second search). Moreover the response time of this request should be **at least 2 minutes**. If it is getting processed before 2 mins then wait for 2 minutes and then return the response. Also return and display the actual metric of the computation (without the 2 min delay) for each of the following computations and as a whole.
    a. How many companies are there in the same country
    b. How many companies with greater diversity are in the same country

c. How much increase/decrease did the company have in its stock price, market share, revenue, expense year by year. (compute all)
d. How many companies have a greater stock price, market share, revenue, expense than this company domestically and globally (compute all)
e. Comment on the growth, stability of the company inferring from the data
f. Predict the next year's stock price, market share, revenue, expense than this company domestically and globally (compute all). Note that in this prediction you should be thinking about how each factor is affected by the other factors

Note: You should only consider the data which is given to you.

An example of Point 3:
User clicks on submit after the initial company search.
a takes 2 secs
b takes 1 sec
…..
f takes 40 secs
total time taken to compute all: 50 secs
But due to the min delay total response time: 2 min

UX Flow [SDE / SDE + Astro Track]

1. User visits the site
2. User logs in / register with 2FA.
3. User forgets password so goes on for password recovery (Branched flow)
4. User logs in and updates the password if required. (Branched flow)
5. User follows the **PSA / PSB** flow and clicks on **SUBMIT**
6. User waits for 2 minutes for the computation to finish
7. User is displayed the information with a good UI based on data visualization (eg. Graphs, Tables etc) (Note that this is applicable for both PSA and PSB)
8. User closes the website or goes on to perform a different computation.
9. User comes back and has an option of history which has the list of all his previous computations. User clicks on it and the page opens instantly

---

Common Points for both PS

Expectations

- Teams are expected to complete the full SDE part of the PS if SDE track is chosen and vice versa for the Astro track.
- Students who have chosen both the tracks are expected to contribute to both parts of the problem statement.

Timeline

Milestones	Timeline
Hackathon PS release + Groups release	Sep 22, 2024
Mid Evaluation (Writeup)	Sep 28, 2024 11:59 PM
End Valuation (PPT + Demo)	Oct 6, 2024 11:59 PM
Release of Results	Oct 9, 2024
Training of Selected candidates for further process starts	Oct 10, 2024

Evaluations

1. The judging will be both on team as well as individual performance.
2. Every contestant has to maintain a sheet of all the individual contributions to the team and will be judged on that.
3. The PS will be broken down into 2 milestones which are MidEvaluation and EndEvaluation where teams have to give a writeup and PPT respectively.
4. The evaluation will be of the form 50 (Mid) + 100 (End) totalling to 150 points in each track. The points of each topic are mentioned above on the PS and you can scale it down for each milestone to know how many points per PS topic you can achieve.
5. Each individual will be given points and for each evaluation 50% of the points are individual basis and 50% are group basis. (For eg. if a milestone is of 100 points, a contestant can get 50 points (max) for his own contributions and 50 points (max) for the team's progress). Ranking will be based on an individual's total score (individual + team).
6. The evaluation will be based on time management, research, progress, uniqueness of the solution and presentation skills.
7. Note that SDE and Astro will have separate leaderboards.

Groups

Every team should make a whatsapp group and fill the form with the link of the whatsapp group for the managers to join. This will be used for solving doubts or if you want any help from the managers. The link of the form is in the links section.

Resources

- Dataset for PSA: PRADAN
  Register to get into the data archive of Chandrayaan-2 payload data. After registering and logging in, go to XSM (X-ray Solar Monitor) where required data and manuals to read and use the data are present. The lightcurve data of XSM is the required input for this analysis.
- References for PSA:
  - Gryciuk et al., Solar Physics, 292, 77, 2017
  - Aschwanden and Freeland, Astrophysical Journal, 754, 112, 2012
- PSB Resource

Submission Links

Milestones	Link	Instructions	Deadline
Q&A 1	Link	Only the team leader should fill it	Sep 24, 2024 3:00 PM
Groups	Link	Only the team leader should fill it	Sep 25, 2024 12:00 PM

Clarifications

1. Students choosing Astro track only are expected to solve **PSA** and will get grace marks to attempt **PSB**. This grace mark totally depends on the effort of the student.
2. **2FA** refers to 2 Factor Authentication (eg: Password + Any other way)
3. **DAU** refers to Daily Active Users
4. If a team has at least one SDE and at least one Astro peep the team is expected to solve both PSA and PSB (Pure-SDE implementation track should not be chosen here)
5. Suppose if a SDE + Astro team only has a SDE submission it will be ranked on the SDE leaderboards but will get 0 points on Astro leaderboards and vice versa.

6. Note that PSB heavily focuses on scalability of your solution also.
7. Everyone should concentrate on the track which they have chosen as the leaderboard rankings will be based on that. Work distribution should be based on the tracks chosen by the members.
8. If a team is doing Pure Astro then they should focus on developing a script to accept the files and perform the computations as described in PSA. For Astro + SDE teams the file should be uploaded from the frontend before the user clicks **SUBMIT**.
9. Note that the entire solution is expected to be hosted.
10. All the teams are expected to provide a **ZERO COST** solution. Zero cost is defined as using services/deployments which are free forever.

References for beginners:

PSB

Note that these resources are just for reference and not for any promotion of any service to be used in the solution in the problem statement.

2FA : https://duo.com/product/multi-factor-authentication-mfa/two-factor-authentication-2fa
Google Login : https://developers.google.com/identity
React Quick Start (Frontend) : https://react.dev/learn
Some good reads:

- https://medium.com/geekculture/system-design-scaling-from-zero-to-millions-of-users-deca270ef784
- https://bytebytego.com/courses/system-design-interview/scale-from-zero-to-millions-of-users
- https://github.com/pragyaasapkota/System-Design-Concepts?tab=readme-ov-file
- https://www.apwide.com/8-deployment-strategies-explained-and-compared/
- https://www.atlassian.com/microservices/microservices-architecture/microservices-vs-monolith
- https://www.locofy.ai/blog/serverful-vs-serverless

Guidelines for MidEvaluation:

Note that the following is the evaluation criteria for the MidEval. The points allotted above in the respective PS will be used for the final EndEval submission. In the MidEval you need to submit the following things:

1. Report (Evaluation metrics of this are described below)
2. Member wise work split (More on this will be given below)
3. Your active Github repo. (Note that this will not be used for evaluation in MidEval)

Reports:

PSA

Write a detailed report which should include the following

- Describe your understanding of the problem statement. [7 Marks]
- What did you understand about solar flares? [7 marks]
- How will you identify the X-ray bursts? [12 marks]
- How will you classify them? [12 marks]
- As per the data provided, how will you extract the features of solar flares? [12 marks]

[Total: 50 Points]

Note that the whole score will be divided by **2** (scale down factor) and awarded to the teams as team scores. (Max Points: 25)

PSB (In General)

Write a detailed report which should include the following

- Detailed architecture of the **whole system.** [30 Points]
- List of services (including the 3rd party services as well as the service(s) you have built) including all the minute details of the plan you are subscribing to. (Note that only ZERO COST free forever solutions will be accepted). Details of the plan should include the following at least (but not limited to) [30 Points]
    - Link of the service
    - Link to the pricing page + Details about which plan you have chosen.
    - Why do you think this is the best fit?
    - Its alternatives ?
- Why do you think your architecture is scalable? [10 Points]
- Till what scale can you accommodate in a zero cost solution? Describe in detail. [10 Points]
- What would have you done differently in your architecture if a zero cost solution was not a requirement? [10 Points]
- What steps did you take to improve your DX and UX ? (DX: Developer Experience, UX: User Experience) [10 Points]

Here DX includes things like tech stack, type safety, code readability, code splitting, CI/CD, DevOps, avoiding vendor locking etc.

PSB Pure SDE Implementation Specific:

- How did you implement the data visualization? [Points: 10]
- How did you store the data and describe the cleanup/computation you have done on the data before storing. Also share the schema of your data. [Points 20]
- What practices did you take to make your system extendible and accept actual real time computations ? [Points: 10]
- How did you do the predictive analysis? [Points: 10]

[Total: 100 + 50 = 150 Points]

Note that the whole score will be divided by **6** (scale down factor) and awarded to the teams as team scores. (Max Points: 25)

For the Individual score the contribution they have made and their impact will be considered. You need to submit a doc in the Google form which will be floated in future. The format for this doc will also be shared by Friday. For the individual the max points a person can get will be 25.

The calculation of the total points will be as follows for MidEval:

[Individual Points] + [Team Points] = 50

25 + 25 = 50