Greetings from ResoluteAI.in!

Mandatory Tasks : Task 1, 2, 3 should be mandatorily completed.

Optional : Any two tasks from 4,5,6,7,8,9,10,11

Data Link : Data set link for all tasks

Role: AI Engineer - Intern

Dataset: PFA

Task 1: Machine Learning - Clustering

Task: Use any clustering technique to extract patterns or segregate data into groups.

User Story: User should be able to provide a data point (a row) and program should be able to identify to which cluster given data point belongs to and why?

Hint: Clustering

Task 2: Machine Learning - Classification

Task: Train any classification model(s) on the train sheet and test your algorithm on the test sheet.

Share target values for the test, we will evaluate from our end.

If using multiple algorithms, plz share target values for each algorithm.

Also share your algorithm's train accuracy and mention the reason behind choosing your algorithm.

Task 3: Python with Basic Streamlit/Flask

Task: Use rawdata as input and derive

1. Datewise total duration for each inside and outside.
2. Datewise number of picking and placing activity done.
3. Represent output of above 3 tasks either in streamlit or flask page.

Sample output shared in output sheet.

Task 4: Big Data Analysis using Pyspark

Task:Explore the shared dataset using pyspark and answer the following questions:

1. Who are the winners of the D1 division in the Germany Football Association (Bundesliga) in the

last decade?

2. Which teams have been relegated in the past 10 years?

3. Does Octoberfest affect the performance of Bundesliga?

4. Which season of Bundesliga was the most competitive in the last decade?

5. What's the best month to watch Bundesliga?

Enter the Spark Environment Of google collab. For your help can refer to the following Link

https://medium.com/analytics-vidhya/getting-started-spark3-0-0-with-google-colab-9796d350d78

Take a screenshot of the Answer and also create a small video explaining your code.

Data is provided check the data folder.

Task 5: Text Analytics using NLP

User Story: User should be able to upload pdf. Application/solution should be able to read the content of pdf and summarize the content into 500 words. Please use “Operations Management.pdf” as input pdf.

Task 6: Q&A using LLMs

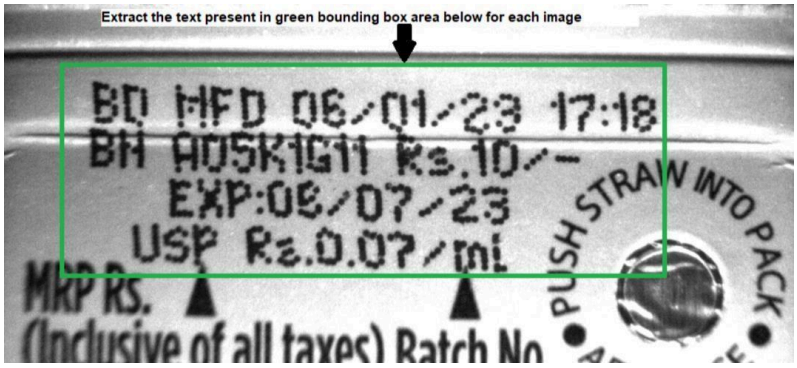
User Story: User should be able to upload a pdf (having textual data) and ask questions, application/solution should provide answers from the context of uploaded pdf. You are free to use enterprise LLM APIs.

Task 7: Computer Vision

Task: Count the number of boxes in the given image.

User story: As a user I should provide a path to the image, and the program should display the total number of boxes present in the image. (you are free to use open-source models and codes, but please ensure that there is no complete copy-paste done)

Task 8: Computer Vision



Task: Read the text as shown in the given image, make sure the code is generalized and work on all the given images in task2 Folder(correct extraction will only be evaluated) .

User story:As a user I should provide a path to the image, and the program should display the **text** present in the image. (you are free to use open-source models and codes, but please ensure that there is no complete copy-paste done)

Task 9: Computer Vision

Task: Count the number of items (products) in the given freezer images **compartment wise**.

Tip: Detect & Locate all the compartments of the Refrigerator using computer-vision techniques, and then count.

User story: As a user I should provide a path to the image, and the program should count and display the information about the number of items/product in the image compartment. (you are free to use open-source models and codes, but please ensure that there is no complete copy paste done).

Task 10: Computer Vision

Task: Fabric defects should be detected using a segmentation approach and localize the defect. Use the given dataset for fabric defect detection.

User story: As a user I should provide a path of the Fabric image, the program should detect the defect and localize it and mask the defective region. (you are free to use open-source models and codes, but please ensure that there is no complete copy paste done)

Task 11: Computer Vision

Task: Detect all the different parts of the clothing, such as sleeves, collar area, neck area, and chest area, in the given dataset for task11.

User story: As a user, I should be able to provide the path of a clothing image. The program should then detect and localize different areas of the clothing, such as sleeves, collar area, neck area, and chest area, by either outlining or masking these regions in the image. (you are free to use open-source models and codes, but please ensure that there is no complete copy paste done))

Important points to consider before submission:

1. You are free to choose open-source models and codes. Please ensure that there is no complete

copy paste done.

2. Please share **only screen recorded videos** and not the code files.

3. Rename your screen recorded file with your name.

4. Videos **without the applicant's name in filename would not be considered for evaluation**.

5. Fill the google form while submitting your assignment, without fail. Google Form Link

6. Need to complete all first three tasks mentioned. And any two tasks from task number 4 to task number 11. However, if stuck at any one task, please move to next and complete as much as possible before the deadline.

7. Approach for all tasks and reading of instructions will be given higher preference over accuracy and code.

Good Luck !!