Mini-Project Guidelines

It has come to our notice that students of CSE, 6th-semester did not fully meet the standards of Mini-Project submission and therefore will be required to re-submit their projects which would be re-evaluated following the guidelines below -

1. All the students who have, in any form, plagiarised the source code, or reports from the web or among themselves have been levied a penalty of **20 marks**. Consequently, the re-submission will only be evaluated against a maximum of **80 marks.** It includes all those -
    a. Who copied the source code from the web or from a fellow student, partially or completely.
    b. Who copied the source code and made changes in variables and function names with no change in business logic whatsoever.
    c. Students whose source code/reports match with other fellow students no matter who did the original work.

2. Plagiarism in future in any form will result in **strict action** and the project will not be evaluated.

3. Please be advised that the entire source code should be committed to **GitHub** and commits should be done incrementally to justify the work done. On the day of evaluation, students may be asked to clone the GitHub repository onto the local machine and run the code for evaluation. You can also deploy the code to a cloud platform like **Heroku** (although not mandatory for evaluation, it will give you an edge during the evaluation process).

4. You should be able to perform the following things in general. Feel free to adapt based on the title of your project -

    a. If you are using already processed data of rather small size, use a database like MySQL, sqlite3, or MongoDB to store the data and then retrieve it using a web portal implemented in a technology stack as per your choice (LAMP, Flask or Django, MEAN etc.). For example, consider a registration portal for students which makes an entry in the database from the registration portal.

    b. You should have a web portal that either fetches the real-time data or utilizes already downloaded data. You are advised to use your own API keys to download the data and build your analysis on top of it.

    c. If you plan on using APIs to download the data using your own API keys, you can process it and store it in CSV files (in case of text data) or any other suitable formats as suits. Use web portals that can take sample data from your file, process it further and store it in the designated format. For example, if you get JSON data, allow a web portal to browse the JSON file and convert it into CSV format or any other required format. In case you have some

unprocessed image that you need to process for further use, provide functionality on a web portal that can browse and process it with the required functionalities.

d. No values should be hard coded as part of the functionality of the project. This means that the program must be shipped as a complete product without any intervention required in the source code whatsoever to alter the course of execution. You should launch the program just once which is expected to accept all the inputs from the user and process accordingly. For example, if you want to insert a new record for your sample testing, the program should have an insert-record functionality that can be accessed from the UI or the console with no manual change allowed in the source code.

e. Students are advised to not take the already labelled data for the training purpose. **Hint-** You can study the way of labelling data and label a few instances on your own (manually or automatically) and then provide analysis to show that the annotation is efficient.

f. You can use various approaches to label the training and test data and mention it clearly in your project report. You can refer to the following paper to familiarise yourselves with the process - Identification of Suicidal Ideation in Tweets

g. In relation to the above point, if you need to choose some features out of many, it should be clearly mentioned as what made you choose some feature over others and show its visualization such as error or accuracy graphs that changes with the addition or deletion of features.

h. For projects on IoT devices, assuming you are not able to procure the required hardware, you can simulate the various functionalities and use them for further analysis. For example, for implementing a smart dustbin system, you can use the programs to generate synthetic data based on some reference studies and use the web portal to show the various related analyses and apply ML and other algorithms if necessary.

i. For the projects that process videos, you should use a web or mobile application to get the input and show output based on the algorithms applied. For example, Detecting sentiments from videos should ask a user for permission to access the device camera and then capture the data for sentiment analysis. There should also be a separate dashboard that shows performance analysis on already downloaded datasets.

j. For projects like Recommendation System, you should create a web portal that is able to recommend other products. If you are using different kinds of recommendation approaches or similarity functions, your web portal should be able to choose one of them and show the required analysis.

k. For a project such as sentiment analysis, you should add the following functionalities-

   i. If you are applying various algorithms or techniques such as stemming or lemmatization, it should clearly show the accuracy metric with or without using these techniques and show the visualization that shows the importance of techniques used.

   ii. You are advised not to use inbuilt algorithms to calculate the sentiments, rather you can apply the dictionary-based approach to calculate and compare the accuracy with inbuilt methods. You can refer following paper to read about this - Sentiment Analysis

5. Project Report Guidelines -
   a. You should include following sections in your report - Abstract, Project Introduction, Methodology which might further include data collection or preparation, data preprocessing and methodology and Result Analysis.
   b. Reports should fully adhere to the report standard. Any form of plagiarism in the report will disqualify the project from re-evaluation.
   c. Refrain from writing too many definitions and formulae in the report. Focus more on the core logic or something that you have implemented to highlight your contribution to the project.
   d. The report should not include more than 8 pages and must be in the following format - Report Format
   e. Include all the references in reference section

6. Feel free to reach out to your project guide for any clarification on the guidelines and understanding of the project.

7. You can use following references for your project -
   a. Twitter Sentiment Analysis
   b. JSON Viewer
   c. Embedding Sentiment Analysis Model into a Web Application
   d. Integrate Machine Learning into Web Applications with Flask
   e. Deploying a Static Web Application to Heroku
   f. Creating REST API Using Flask
   g. Hotel Recommendation System
   h. Sarcasm Detection-Using Hybrid NN