Data Science Mini Project Guidelines

Date to Select Topic: 02-07-2025 (EOD)

Dear Students,

You are required to complete a **mini-project** as part of your course. Please follow the guidelines below carefully:

1. Project Topic Selection

- You can **choose any one topic** from the list given below:
    - **Credit Card Default Prediction: Dataset link**
    - **Dry Bean Classification: Dataset link**
    - **Mushroom Classification: Dataset link**
    - **Term Deposit Prediction Model (Bank Marketing): Dataset link**
    - **Steel Industry Energy Consumption: Dataset link**
    - **Bike Sharing: Dataset link**

Important:

- Each student must select a **unique topic**.
- **No two students are allowed to work on the same topic.**

---

2. Topic Confirmation

- After selecting your topic, **fill out the Google Sheet** shared with you to confirm your selection.
- This will help us track and avoid topic duplication.
- Deadline to select and update your topic: **02-07-2025 EOD**

3. Project Implementation

- Use **Google Colab** to complete your project.
- Download the dataset directly from the links provided.
- Structure your notebook professionally with clear formatting.

Tip: You can use **HTML tags** in Colab (e.g., <h1>, <h2>, <p>) to create neat headings and subheadings.

4. Roadmap to Follow

Refer to the shared roadmap for guidance, but your project should follow this structure:

5. Submission Format

- Ensure all headings are properly aligned.
- Use comments, markdown cells, and visualizations clearly.
- Keep your code clean and readable.

6. GitHub Submission Guidelines (Final Step – Mandatory)

You must **submit your final project on GitHub**. A Google Form to submit your **GitHub repository link** will be shared with you later. Make sure your GitHub repository is organized and contains all the required components.

Your GitHub Repository Must Include the Following Files:

1. **Final Jupyter/Colab Notebook** (`.ipynb`)
    - This should contain your complete project implementation, EDA, model building, evaluation, and conclusions.

2. **Dataset File** (`.csv`)
    - Upload the raw or cleaned version of the dataset used in the project.
3. **Saved Model File**
    - Save your best-performing model using appropriate methods (`joblib`, `pickle`, or `model.save()` for deep learning) and upload it.
4. **`README.md` File**
   This file should clearly describe your project and must include the following sections: Project Title, Objective, Dataset Used(Name and source of the dataset, Link to the original source), Models Used(Mention the machine learning models applied, Briefly state why you chose them), Key Results(Include final evaluation metrics (accuracy, F1 score, AUC, etc.), State which model performed best and the outcome

Notes:

- Keep your GitHub repository **clean and professional**.
- Use proper file names and folders if needed.
- Make sure the notebook is well-formatted and all cells are run before submission.

Best wishes,
Greeshma S
Data Science Mentor