Protect Lab Internship Task: Employee Disengagement Prediction

---

Problem Statement:

Title: Predicting Employee Disengagement in Remote Work Settings

Context: You have been hired as a data scientist at a fictional remote-first company, **Protect Inc.**. The HR department has noticed that some employees gradually disengage from work before leaving the organization. They want you to use behavioral data to identify early signs of disengagement.

Your task is to analyze a time-series dataset of employee activity over 6 weeks and build a machine learning model to predict whether an employee will become disengaged.

Dataset Information:

File Name: `employee_behavior_timeseries.csv`

Dataset Description: The dataset includes weekly behavioral metrics for 100 employees over 6 weeks. Each employee has 6 rows (one per week).

Column Name	Description
employee_id	Unique identifier for each employee
week_number	Week index (1 to 6)
average_meeting_hours	Average hours spent in meetings during the week
code_commits_per_week	Number of code commits or completed work items during the week
messages_sent_slack	Messages sent on the company's Slack/Chat platform
screen_active_minutes	Screen time in minutes during work hours
sentiment_score	Weekly sentiment score from internal feedback (ranging from -1 to +1)
timezone_alignment	1 if employee works in same timezone as their team, 0 otherwise
disengaged	Target label: 1 = disengaged in the last 2 weeks, 0 = otherwise

---

Task Instructions:

1. **Understand and Explore** the dataset:
    - Perform EDA to understand the trends and distributions

    - Visualize behavioral differences between engaged and disengaged employees
2. **Build Features and Train a Model**:
    - You may aggregate weekly data per employee (mean, min, max, etc.)
    - Apply high accuracy classification
3. **Explainability & Ethics**:
    - Identify top contributing features
    - Discuss any ethical concerns in using behavioral monitoring for HR decisions

---

Post-Task Reflection Questions (To Be Answered After Modeling):

Each candidate must answer the following in the video and as a short PDF/text response:

1. **What were the top 3 features contributing to disengagement predictions, and why?**
2. **How did you process the time-series nature of the data (i.e., per week entries)?**
3. **What patterns did you notice among disengaged employees?**
4. **How would you ethically use this model in a real-world HR setting?**
5. **What improvements would you suggest if you had access to more data?**

Deliverables (What to Submit):

1. 3-Minute Video Presentation (Required) - Introduce the problem (briefly) - Show your EDA and insights - Explain your model and performance metrics - Discuss feature importance and key takeaways - Answer the reflection questions

2. Clean Code Notebook - Well-commented Python notebook (.ipynb or Colab link) - Must run without errors

3. File Naming Convention: - Video: `YourName_ProtectTask.mp4` - Notebook: `YourName_ProtectModel.ipynb` - Optional response PDF: `YourName_ProtectAnswers.pdf`

4. Submission Method: - Submit via Google Form and also upload to the class room

5. Deadline: - [28 June, 2025, 04:00 PM]
