Data Science for Entrepreneurs 2025/6- Final Project

The purpose of the final project is to apply the data science concepts, tools, and methodologies learned throughout the course to a **real-world, product-oriented business problem**. You will work in teams of 2-3 students to identify a challenge, build and evaluate a model, and deploy a working solution. Each team member is expected to contribute meaningfully to the project. Be prepared to explain and justify your design choices.

Please note: teams of 4 students are permitted only if one team member is a reserve duty soldier.

Problem Selection & Data Source

Your first task is to identify a product-related or business-oriented challenge that can be addressed using a data-driven approach.

You can use one or more relevant datasets. A single dataset may support different analyses, so clearly define your problem. Datasets for your project must be from the following public repositories:

- Kaggle: https://www.kaggle.com/datasets
- UCI Machine Learning Repository: https://archive.ics.uci.edu/datasets

Before proceeding, your team must update this form with student information, a link to the dataset, and a **short description** of the business problem and the type of model you plan to build. Approval is required before implementation.

Part 1 - Project Implementation (40%)

Once your proposal is approved, you will implement a complete data science pipeline. Your project should clearly demonstrate the methodologies and best practices covered in the course. Your submission must include the following components:

1. **Problem Definition:** Clearly state the business problem, stakeholders, and how a machine learning model will address the challenge.
2. **Data Analysis and Preparation:** Explore and prepare the data thoroughly for modeling, showing strong understanding and thoughtful preprocessing.

3. **Model Building:** Choose and implement a suitable algorithm discussed in class and justify your choice. Compare models if needed, but explain your final selection clearly with well-structured, commented, reproducible code.
4. **Evaluation & Insights:** Evaluate model performance, define success, pick suitable metrics, validate results, and highlight key insights specific to your data and model while avoiding generic statements.
5. **Deployment:** Use **Gradio** to create an interactive app from your Python notebook allowing users to input features and receive model predictions, demonstrating practical use of your solution.

Part 2 - Oral Presentation Guidelines (60%)

- **App Demo on January 13:**
  Live demonstration of your application, up to 7 minutes, including two use cases.
- **Q&A:**
  2 questions for each team member. All members must actively participate and demonstrate understanding of the project.

Submission Requirements

Each team must submit:

1. **Complete Python code** including data processing, model training, and Gradio deployment.
2. **A summary document (Word or PDF) up to two pages in length** that describes the problem and motivation, data, methodology, results, application design choices and key insights and conclusions. Make sure to include 2 graphs that best visualize your key findings.

If you use terminology, models, or techniques not covered in class, you must include appropriate references.

Good luck!