Project 01: AI-Powered Loan Default Prediction System for Banking and Fintech

AI Project Management Lifecycle

1. Project Initiation and Business Understanding
2. Data Strategy and collection
3. Data Preprocessing and Feature Engineering
4. Model Selection and Development
5. Model Training, Evaluation, and Validation
6. Deployment and Integration
7. Monitoring and Maintenance

Week 01 Tasks

Write a report on :

- Project Initiation and Business Understanding
- Data Strategy and collection

For the report, follow the sample report mentioned below.

Name: Faiyaz bin Zaman

E1 Member ID: E1-24241281-MHTLM5HL

Wings:

Wing 1 (Intelligence & Perception)

Wing 2 (Modeling & Insight)

Wing 3 (Systems & Applications),

Project 01: AI-Powered Loan Default Prediction System

Project Initiation and Business Understanding

Problem Definition: Banks have a major problem, they don't know for sure which customers will pay back a loan and which ones won't. This uncertainty causes them to lose a lot of money on bad loans or defaults. Currently, checking every application by hand is too slow and leads to mistakes. **The Solution is,** we will build an AI system that looks at a loan application and predicts a simple Yes or No, is this person likely to default?

Business Objective:

- **Main Goal:** Stop losing money. The AI will recognize high risk applicants before the bank approves the money.
- **Secondary Goal:** Speed up service. Low risk customers get approved instantly
- **The Result:** The bank saves millions in lost funds, and customers get faster answers.

Feasibility Assessment:

- **Technical, High:** This is a solvable problem. Standard AI models like Random forest or xgboost are famous for being good at this specific task.
- **Data, High:** Banks already own the data we need such as transaction history, income proof.
- **Economic, High:** The cost of building this software is quite small compared to the money saved by preventing a few large defaults.

Data Strategy and Collection

We will follow the **Epoch One Data Ethics Guidelines** to ensure our data handling is legal.

Real World Data Collection Strategy: For this assignment, we are using a Kaggle file. However, if this were a real Wing 3 project at a bank, we would not just "download a file." We would build a pipeline to pull data from three live sources:

- **Application Data:** Information the user types in Salary, Job Title or Loan Purpose.
- **Internal History:** Data the bank already has
- **Credit Reports:** We would connect to a Credit Bureau like Equifax via api to get their credit score.

Data Ethics & Safety: Public datasets can be unsafe or biased. Based on the **Epoch One Data Ethics Guideline**, here is how we will handle it:

- **Privacy (Guideline 3: Data Minimization):**
    - **The Rule:** Collect only what is necessary.
    - **The Fix:** We will scan the dataset for "Personally Identifiable Information" (PII) like Names, Phone Numbers, or National IDs. These will be **deleted** immediately. If we need to track rows, we will "hash" (scramble) the IDs so they cannot be traced back to a real person (Pseudonymization).
- **Fairness (Guideline 6: Fairness & Bias Prevention):**
    - **The Rule:** Avoid biased sampling or exclusion.
    - **The Fix:** Historical data often contains human bias such as past loan officers rejecting people based on gender or location. We will remove sensitive columns like Gender, Race, or Religion from the training data so the AI cannot learn to discriminate based on them. We will also check that our dataset isn't just from one specific group of people to ensure it represents everyone fairly.

Dataset: Loan Default Dataset [Kaggle]