

Milestone 1 Summary – Requirements & Data Collection

1. Objective

The objective of Milestone 1 is to clearly define the project requirements, understand the problem domain, and prepare all essential datasets and KPIs needed for building an AI-driven loan recovery prediction system. This milestone ensures a strong foundation before any data engineering or modelling work begins.

2. Purpose

Milestone 1 focuses on establishing the **functional and technical direction** for the CreditPathAI project.

The purpose is to:

- Understand the business problem of loan default prediction
- Identify key project components such as ingestion, modelling, and recommendation
- Collect and arrange the datasets required for analysis
- Document all KPIs that the system will track
- Set up the GitHub repository for version-controlled development

This milestone defines **what the system will achieve, how it will work, and what data it needs.**

3. Work Performed

3.1 Project Understanding & Requirement Finalization

Using the project description in the PDF, the scope of CreditPathAI was finalized:

- Predict borrower default risk
- Model repayment behaviour
- Recommend personalized recovery strategies
- Support a full ML workflow from ingestion → modelling → API → dashboard

The architecture, milestones, and tech stack were reviewed and finalized.

3.2 KPI Documentation

From the dataset and the project objective, the following KPIs were defined:

Business KPIs

- Total Borrowers
- Total Loan Amount
- Average Loan Amount
- Default Rate (%)
- Average Credit Score
- Average Interest Rate
- Average DTI Ratio
- Average Months Employed

These KPIs help understand the borrower profile and credit risk distribution.

Model KPIs (for later milestones)

- AUC-ROC
- Precision, Recall, F1-Score
- Accuracy
- Confusion Matrix (TP, FP, TN, FN)

Data KPIs

- Data ingestion speed
- Missing value summary
- Duplicate count
- Memory usage

These KPIs were documented as required deliverables for Milestone 1.

3.3 Dataset Collection & Preparation

As mentioned in the PDF:

- Kaggle Loan Default Dataset
- Microsoft R Server Loan Credit Risk Dataset

You prepared a unified dataset containing fields such as:

LoanID, Age, Income, LoanAmount, CreditScore,
MonthsEmployed, NumCreditLines, InterestRate,
LoanTerm, DTIRatio, Education, EmploymentType,
MaritalStatus, HasMortgage, HasDependents,
LoanPurpose, HasCoSigner, Default

This dataset fully matches the project schema required for further modelling.

3.4 GitHub Project Setup

A GitHub repository was prepared (as required by PDF for Milestone 1):

- Created directory structure
- Added README base
- Pushed initial dataset and scripts folder

This ensures smooth progress across further milestones.

4. Conclusion

Milestone 1 establishes the project's foundation by finalizing requirements, identifying KPIs, collecting datasets, and setting up the development environment. All objectives stated in the project PDF have been successfully completed. The collected dataset and documented KPIs will serve as the basis for data ingestion, exploration, feature engineering, and modelling in upcoming milestones.