

## Data Collection and Preprocessing Phase

### Data Collection Plan & Raw Data Sources Identification Report:

The key raw dataset was obtained from Kaggle's "Loan Prediction Problem" competition archive. A sample is provided as **loan\_prediction.csv** (614 rows  $\times$  13 cols).

### Data Collection Plan:

| Section                        | Description   |
|--------------------------------|---|
| Project Overview               | We aim to build a machine-learning model that predicts whether a loan application will be approved, based on demographic, financial, and credit attributes supplied by applicants.  |
| Data Collection Plan           | <ul style="list-style-type: none"><li>● Prioritise public datasets with rich applicant-level features (gender, income, credit history, etc.).</li><li>• Source primary loan-approval data from reputable repositories (Kaggle/UCI).</li><li>• Acquire supplementary credit-bureau information if required.</li><li>• Ensure licensing permits academic use.</li><li>• Validate schema consistency across sources prior to merge..</li></ul> |
| Raw Data Sources<br>Identified | The key raw dataset was obtained from Kaggle’s “Loan Prediction Problem” competition archive. A sample is provided as <b>loan_prediction.csv</b> (614 rows × 13 cols).  |

|  |   |
|--|---|
|  | marital status, income, and loan-related details for machine learning analysis. |
|--|---|

### Raw Data Sources Report:

| Source Name    | Description  | Location/URL  | Format | Size  | Access Permissions |
|----------------|--|---|--------|-------|--------------------|
| Kaggle Dataset | Historic bank-loan records with demographics, incomes, loan terms, and target label Loan_Status. | <a href="https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset">https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset</a> | CSV    | 15 kB | Public             |
| UCI            | This data concerns credit card applications; a good mix of attributes                            | <a href="https://archive.ics.uci.edu/dataset/27/credit+approval">https://archive.ics.uci.edu/dataset/27/credit+approval</a>   | CSV    | 14 kB | Public             |