

## Data Collection and Preprocessing Phase

<b>Date</b>	11 March 2025
<b>Team ID</b>	740052
<b>Project Title</b>	AI-Based Intelligent Insight Extractor
<b>Maximum Marks</b>	2 Marks

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

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

### Data Collection Plan Template

Section	Description
Project Overview	The AI-Based Intelligent Insight Extractor is developed to automate the detection and resolution of data quality issues in the loan approval prediction system. It leverages AI-driven modules to profile data, analyze documentation, recommend preprocessing strategies, and provide actionable insights to improve model training efficiency and performance.
Insight Extraction Plan	The system will ingest raw datasets from internal banking sources and external repositories. It will perform automatic data profiling, detect missing or inconsistent values, classify data types, and suggest remediation strategies (e.g., imputation, encoding, scaling). The insights are integrated with model preparation stages for streamlined workflow automation.

## Raw Data Sources Identification Template:

Field	Description
<b>Source Name:</b>	Unique name or ID for the data source
<b>Source Type:</b>	Structured (e.g., SQL), Semi-structured (e.g., JSON, XML), Unstructured (e.g., PDFs, images, audio)
<b>Description:</b>	Brief overview of the data source and its relevance
<b>Owner / Provider:</b>	Internal team, external API, data vendor, etc.
<b>Update Frequency:</b>	Real-time, hourly, daily, weekly, etc.
<b>Format:</b>	CSV, JSON, XML, PDF, text, audio, video, etc.