

Data Collection and Preprocessing Phase

Date	18 June 2025
Team ID	SWTID1749880888
Project Title	Prosperity Prognosticator: Machine Learning for Startup Success Prediction
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Columns with date/time data (e.g., first_funding_at, closed_at)	Moderate	Dropped or transformed to numerical formats (e.g., age in years)
Kaggle Dataset	Mismatch in input features between training and prediction models	High	Re-trained model using only 7 essential features consistently
Kaggle dataset	Categorical data in state_code, category_code, has_roundB, etc.	Moderate	Applied Label Encoding to convert categorical to numeric