

Data Collection and Preprocessing Phase

Date	09 July 2024
Team ID	SWTID1720104839
Project Title	Human Resource Management: Predicting Employee Promotions Using Machine Learning
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	Null values in education and previous year rating	High	Replacing values using mode for education and previous year rating
Kaggle	Imbalanced data as there are more number of 0's compared to 1's in label is promoted	High	Using over sampling to increase the number of 1's so both 1 and 0 are balanced
Kaggle	Categorical attributes- education and department	Medium	Using label encoder

Kaggle	Outliers present in length of service	Medium	Handling using capping method
--------	---------------------------------------	--------	-------------------------------