

Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 14 July 2024 |
| Team ID | SWTID1720174920 |
| 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 |
|---------------------------|--|----------|---|
| Employee Records | Missing values in key fields | High | Implement data imputation techniques such as mean/mode imputation for numerical/categorical data or use more sophisticated methods like KNN imputation. |
| Historical Promotion Data | Outliers and anomalies in promotion records (e.g., unusually rapid promotions) | High | Identify and investigate outliers using statistical methods or domain knowledge. Remove or correct anomalies based on consistency checks and validation with HR experts. Apply robust outlier |

| | | | |
|---------------------------|--|----------|--|
| | | | detection algorithms like IQR or Z-score methods |
| Employee Surveys | Duplicate survey responses | Moderate | Use data deduplication techniques such as identifying and removing duplicate rows based on unique employee identifiers and timestamp analysis. Ensure each employee has only one survey response. |
| Skills and Certifications | Outdated or unverified skills and certification data | Moderate | Regularly update skills and certification records through self-reporting mechanisms and validation with professional certification bodies. Implement periodic audits and verification processes to ensure data accuracy. |
| Job History | Inconsistent job title naming conventions | Moderate | Standardize job titles using a predefined job title taxonomy. Implement a mapping system to convert all job titles to a consistent format. |