**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | xxxxxx |
| 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  Dataset | Missing values in the ‘education’ and ‘previous year rating’ columns | Moderate | Use mean/median imputation |
| Kaggle Dataset | Categorical data in the dataset | Moderate | Encoding has to be done in the data |
| Kaggle Dataset | Negative Data in the dataset | Low | List comprehension used to find the negative data the dropped negative data |
| Kaggle Dataset | Imbalanced Data | High | To handle imbalanced data resampling are done with SMOTE |