Data Collection and Preprocessing Phase

Date	25 June 2025
Team ID	SWTID1750155746
Project Title	Human Resource Management: Predicting Employee Promotions using Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

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:

Section	Description
Project Overview	The machine learning project aims to predict employee promotion eligibility based on individual attributes. Using a dataset with features such as department, education, training score, and performance history, the objective is to build a model that accurately classifies whether an employee is eligible for promotion, enabling fair and data-driven decision-making in the HR process.
Data Collection Plan	- Search for datasets related to employee promotions, performance metrics, and HR records. - Prioritize datasets with diverse roles, departments, and evaluation parameters.
Raw Data Sources Identified	The raw data source for this project is a dataset obtained from Kaggle, a well-known platform for data science datasets. The dataset contains employee-related information including department, education, number of trainings, age, previous year rating, length of service, KPIs met, awards won, and average training score — all of which are used

	for machine learning analysis to predict promotion eligibility.				

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	employee_id, department, region, education, gender, recruitment_channel, no_of_trainings, age, previous_year_rating, length_of_service, KPIs_met >80%, awards_won?, avg_training_score, and the target variable is_promoted.	https://drive.google.com/file/d/14eQR1VWHwuomPaXdKuIkZEpSstvav8Db/view	CSV	3.6 Mb	Public