

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 Collection Plan & Raw Data Sources Identification Template

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 Template

Section	Description
Project Overview	Brief overview of the machine learning project and its objectives.
Data Collection Plan	Mention from which sources the data are going to be collected.
Raw Data Sources Identified	List the raw data sources with relevant details (as a short description).

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
-------------	-------------	--------------	--------	------	--------------------

Dataset 1	<p>This dataset contains information about employee promotions. Each row represents an individual employee with various attributes that may influence their promotion status. Below is a description of each column in the dataset:</p> <ol style="list-style-type: none"> employee_id: A unique identifier for each employee. department: The department where the employee works (e.g., Sales & Marketing, Operations, Technology) region: The geographical region where the employee is located (e.g., region_7, region_22). education: The highest level of education attained by 	https://drive.google.com/file/d/14eQR1VWHwuomPaXdKuIkZEpSstvav8Db/view	CSV	3.58MB	Public
-----------	---	---	-----	--------	--------

	<p>the employee (e.g., Master's & above, Bachelor's).</p> <p>5. gender: The gender of the employee (e.g., f for female, m for male).</p> <p>6. recruitment_channel: The channel through which the employee was recruited (e.g., sourcing, other).</p> <p>7. no_of_trainings: The number of training programs the employee has completed.</p> <p>8. age: The age of the employee.</p> <p>9. previous_year_rating: The performance rating of the employee in the previous year, on a scale from 1 to 5.</p>				
--	--	--	--	--	--

	<p>10. length_of_service: The number of years the employee has been with the company.</p> <p>11. KPIs_met >80%: Indicates whether the employee met more than 80% of their Key Performance Indicators (KPIs) (1 for yes, 0 for no).</p> <p>12. awards_won?: Indicates whether the employee has won any awards (1 for yes, 0 for no).</p> <p>13. avg_training_score: The average score of the employee in training programs.</p> <p>14. is_promoted: Indicates whether the employee was promoted (1 for yes, 0 for no).</p>				
--	---	--	--	--	--

	<p>This dataset can be used to analyze the factors influencing employee promotions and build predictive models to identify potential candidates for promotion based on their attributes.</p>				
--	--	--	--	--	--