

Data Collection and Preprocessing Phase

Date	09 July 2024
Team ID	SWTID1720455879
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	The machine learning project aims to build a employee promotion prediction based on the employee information.Using the dataset with features department,education,age,KPI,awars won,average level of training,previous year rating and some other.The objective is to build a model that accurately classifies if the employee is eligible for promotion(1) or not(0).
Data Collection Plan	Searching for datasets in kaggle and uci websites and identifying the most relevant dataset based on the features present and size of the dataset and selecting that dataset.

Raw Data Sources Identified	<p>Kaggle: Kaggle is a popular platform for datasets. You can search for terms like "employee promotion", "HR analytics", or "employee performance".</p> <p>UCI Machine Learning Repository: They provide various datasets across different domains, including employee-related data.</p> <p>Government Databases: Some government agencies provide datasets related to workforce statistics and employment trends.</p>
-----------------------------	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	<p>Multiple attributes have been provided around Employee's past and current performance along with demographics.</p> <p>Features:</p> <ul style="list-style-type: none"> • employee_id • department • region • education • gender • recruitment_channel 	<p>https://www.kaggle.com/datasets/arashnic/hr-ana</p>	CSV	935kB	Public

	<ul style="list-style-type: none"> • no_ of_ trainings • age: Age of Employee • previous_ year_ rating • length_ of_ service • awards_ won? • avg_ training_ score • is_promoted 				
--	---	--	--	--	--