Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	**740071**
Project Title	Work Force Retention System
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	A workforce retention system using machine learning involves developing an algorithm that can predict the likelihood of an employee leaving the company. The project aims to improve employee retention rates and reduce the costs associated with high turnover rates. The workforce retention system can also provide actionable insights to help managers and HR professionals develop targeted retention strategies. For example, if the model identA workforce retention system using machine learning involves developing an algorithm that can predict the likelihood of an employee leaving the company. Overall, a workforce retention system using machine learning can help companies improve employee retention rates, reduce costs associated with turnover, and create a more engaged and satisfied workforce.

Data Collection Plan	• Search for datasets related to Work Force, Employee Turnover, and More details. • Prioritize datasets with diverse demographic information.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle & UCI, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as gender,

	marital status, income, and loan-related details for machine learning analysis.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details (gender, marital status), financial metrics (income, loan amount), and loan approval outcomes.	https://www.kaggle.com/datasets/rishikeshkonapure/home-loan-approval?select=loan_sanction_train.csv	CSV	15 kB	Public
UCI	This data concerns credit card applications; a good mix of attributes	https://archive.ics.uci.edu/dataset/27/credit+approval	CSV	13.6 kB	Public