

Data Collection and Preprocessing Phase

| | |
|---------------|---|
| Date | 19 July 2025 |
| | |
| ProjectName | Machine Learning Approach for Employee Performance Prediction |
| Maximum 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 | This project collects diverse employee data—roles, experience, education, and performance metrics—to build a predictive model that helps organizations make informed decisions on talent management and workforce optimization. |
| Data Collection Plan | <ul style="list-style-type: none"> Identify and gather datasets containing employee profiles, performance metrics, and organizational context. Ensure diversity in roles, departments, and demographics. |
| Raw Data Sources Identified | The raw data sources for this project include dataset obtained from Kaggle, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as productivity, incentive for machine learning analysis. |

Raw Data Sources Report:

| Source Name | Description | Location/ URL | Format | Size | Access Permissions |
|----------------|--|---|--------|------|--------------------|
| Kaggle Dataset | The dataset comprises employee details(team no, incentive), Performance metrics(smv, idle time,), and productivity outcome | https://www.kaggle.com/datasets/utkarshsarbah/productivity-prediction-of-garment-employees | CSV | 95KB | Public |