|  |  |
| --- | --- |
| **Date** | 31 July 2025 |
| **Project Name** | Employee Performance Prediction |

**Data Collection Plan Project Overview:**

This machine learning project aims to predict employee productivity based on various employee attributes such as department, experience, education, working hours, and other relevant behavioral and demographic features. The goal is to build a model that classifies or predicts productivity scores, enabling better resource allocation, workforce planning, and performance improvement strategies.

**Data Collection Plan:**

* Search for datasets related to employee performance, workplace productivity, and human resource analytics.
* Prioritize datasets that include a mix of demographic, job-role, and productivityrelated variables.
* Ensure data covers diverse roles and departments for broader model generalization.
* Use public repositories with open access for ease of reproducibility.

**Raw Data Sources Identified**

The raw data sources for this project were obtained from open repositories like **Kaggle**, which provide high-quality and well-labeled datasets suitable for supervised learning tasks.

**Raw Data Sources Report**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source**  **Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access**  **Permissions** |
| Kaggle  Dataset | Dataset includes features like department, education, experience, working hours, and productivity scores of employees from various companies. | [https://www.kaggle.com](https://www.kaggle.com/) /datasets/utkarshsarbahi/  productivity-predictionof-garment-employees | CSV | ~93 kB | Public |