

Data Collection and Preprocessing Phase

Date	11 July 2024
Team ID	SWTID1720178802
Project Title	Garment worker productivity prediction
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	This project aims to develop a predictive model for garment worker productivity using machine learning algorithms. By collecting and preprocessing historical production data, selecting relevant features, and training regression models, the model will predict productivity levels. The best-performing model will be optimized and deployed for real-time use, helping improve operational efficiency and employee management.
Data Collection Plan	<ul style="list-style-type: none"> • Search for datasets related to garment production workers, • Gather historical data from production systems, HR records, and operational logs.
Raw Data Sources Identified	The raw data source for this project is a dataset obtained from Kaggle, the popular platform for data science competitions and repositories.

--	--

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	This dataset includes important attributes of the garment manufacturing industry such as date, quarter, department, day, team, targeted productivity, SMV.	https://www.kaggle.com/datasets/isahadss/productivity-prediction-of-garment-employees	CSV	16 kB	Public