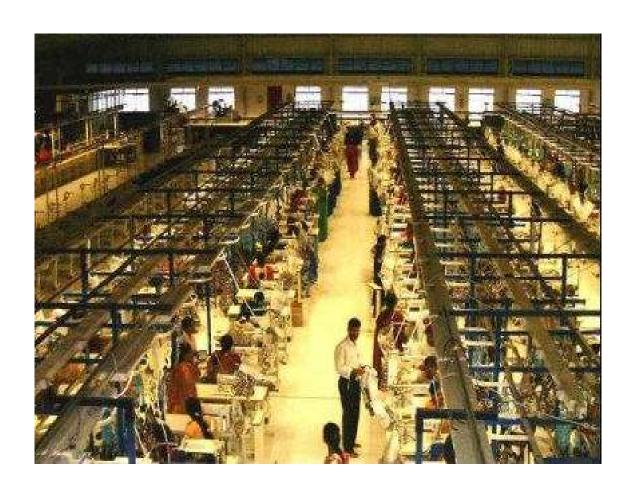# PROJECT

A Supervised Learning on Productivity Prediction of Garment Employees

About Data

The dataset includes important attributes of the garment manufacturing process and the productivity of the employees which had been collected manually and also been validated by the industry experts. Repository located at the following URL:

https://www.kaggle.com/ishadss/productivity-prediction-of-garment-employees.There are **1197** observations and **15** Variables in the Data Set.

Intention

- To build a predictive model and predict the productivity of garment employees.
- To understand the features and find which play a key role in increasing productivity.

Description of the features collected

01 date: Date in MM-DD-YYYY

02 day: Day of the Week

03 quarter: A portion of the month. A month was divided into four quarters

04 department: Associated department with the instance

05 team no: Associated team number with the instance

06 no of workers: Number of workers in each team

07 no of stylechange: Number of changes in the style of a particular product

08 targeted productivity: Targeted productivity set by the Authority for each team for each day.

09 smv: Standard Minute Value, it is the allocated time for a task

10 wip : Work in progress. Includes the number of unfinished items for products

11 overtime : Represents the amount of overtime by each team in minutes

12 incentive : Represents the amount of financial incentive (in BDT) that enables or motivates a particular course of action.

13 idletime : The amount of time when the production was interrupted due to several reasons

14 idlemen : The number of workers who were idle due to production interruption

15 actual_productivity : The actual % of productivity that was delivered by the workers. It ranges from 0-1.

This dataset can be used for regression purpose by predicting the actual_productivity range (0 -1) or for classification purpose by transforming the actual_productivity into different classes (0,1).

Evaluation Metric

The evaluation metric used is root mean squared error (RMSE)

Steps followed to obtain the final model

1. Importing Libraries

2. Data cleaning and EDA

3. Handling unknown

4. Model building

   4.1 Selection of Dependent and Independent variables

   4.2 Split data for train and test

   4.3 Machine Learning Algorithm selection

      The following machine learning supervised algorithms are performed on the dataset and have obtained the RMSE value.

      Linear Regression : (RMSE: 0.14079529970226956)

      Ridge Regression : (RMSE: 0.14078851341948653)

      Lasso Regression : (RMSE: 0.14041619836062005)

      ElasticNet Regression : (RMSE: 0.1405798821432438)

      RandomForest Regression : (RMSE: 0.12647794294604256)

      XGBoost : (RMSE: 0.14409171411183166)

      RandomForest Regression Model has the least Root Mean Squared Error, so we select RandomForest Model as our final model which is robust.

      Performed hyper parameter tuning using GridSearchCV and the best parameter selected :Criterion='mse', max_depth:=10, n_estimators=100, random_state=42

RESULT:

Final model is Random Forest model and it is saved as a pickle file for further use.

RMSE: 0.12647794294604256