

Model Development Phase Template

Date	11 July 2024
Team ID	SWTID1720178802
Project Title	Garment worker productivity prediction
Maximum Marks	6 Marks

Model Selection Report

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

Model Selection Report:

Model	Description	Performance Metric (e.g., Accuracy, F1 Score)
LinearRegression	Linear Regression fits a linear model with coefficients $w = (w_1, \dots, w_p)$ to minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation.	Mean Squared Error: 0.01407603 R2 Score: 0.255140
Random Forest Regressor	A random forest is a meta estimator that fits a number of decision tree regressors on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.	Mean Squared Error: 0.00856469 R2 Score: 0.546783

Gradient Boosting Regressor	This estimator builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage a regression tree is fit on the negative gradient of the given loss function.	Mean Squared Error: 0.009747806 R2 Score: 0.4841768
XGB Regressor	It is a machine learning model from the XGBoost library for regression tasks. It uses gradient boosting with decision trees to predict continuous values, optimizing performance through techniques like regularization and parallel processing.	Mean Squared Error: 0.00960215 R2 Score: 0.4918842
Ridge	This model solves a regression model where the loss function is the linear least squares function and regularization is given by the l2-norm. Also known as Ridge Regression or Tikhonov regularization. This estimator has built-in support for multi-variate regression	Mean Squared Error: 0.014047249 R2 Score: 0.256663825
Lasso	It is a type of linear regression that includes a regularization term in the model. This regularization term (L1 penalty) forces the absolute size of the regression coefficients to be minimized, effectively shrinking some coefficients to zero, thereby performing variable selection and regularization simultaneously. This helps in creating simpler and more interpretable models, especially useful when dealing with high-dimensional data.	Mean Squared Error: 0.019020639 R2 Score: -0.00651227