Model Optimization and Tuning Phase Template

Date	7th July 2024
Team ID	739719
Project Title	Garment Workers Productivity Predictions
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

Optimized models using GridSearchCV, selecting XGBoost Regressor for its superior performance. Trained and validated the final model on the full dataset, ensuring accurate productivity predictions.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Model 1 Linear Regression	**Hyperparameters:** **fit_intercept:** Whether to calculate the intercept for this model (True or False). **Normalize**:Whether to normalize the input variables(True or False)	fit_intercept=True, normalize=False
Model 2 Random Forest	**Hyperparameters:** **n_estimators:**The number of trees in the forest. **max_depth:**The maximum depth of the tree. **min_samples_split:**The minimum number of samples rerquired to split an internal node.	n_estimators=200, max_depth=20, min_samples_split=5

Model 3 XGBoost Regressor	- **n_estimators:** The number of boosting rounds. - **max_depth:** The maximum depth of a tree. - **learning_rate:** Step size shrinkage used in update to prevent overfitting.	n_estimators=300, max_depth=6, learning_rate=0.1, subsample=0.8, colsample_bytree=0.8

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Model 1 XGBoost	**Baseline Metric:** The initial performance metric value for XGBoost before optimization is 0.015	Optimized Metric: The performance metric value for XGBoost after hyperparameter tuning and optimization is 0.123
Model 2 Gradient Boost	**Baseline Metric:** The initial performance metric value for Gradient Boost before optimization is 0.012	**Optimized Metric:** The performance metric value for Gradient Boost after hyperparameter tuning and optimization is 0.116

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Model 1 XGBoost Regressor	The XGBoost Regressor was chosen as the final optimized model because it consistently delivered superior performance metrics during the evaluation phase. Its advanced capabilities in handling complex and high-dimensional data, along with its built-in regularization mechanisms, made it the best candidate for accurately predicting garment workers' productivity.