Model Optimization and Tuning Phase Template

Date	15 July 2024
Team ID	SWTID1720108643
Project Title	Garment Worker Predictivity Prediction
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Linear Regressor	-	-

Decision Tree Regressor	`dt = DecisionTreeRegressor()`
	
	`param_grid = {`
	`'criterion': ['squared_error', 'friedman_mse',`
	`'poisson', 'absolute_error'],`
	`'splitter': ['best', 'random'],`
	`'max_depth': [None, 10, 20, 30],`
	`'min_samples_split': [2, 5, 10],`
	`'min_samples_leaf': [1, 2, 4],`
	`'max_features': [None, 'sqrt', 'log2', 0.5, 1],`
	`}`	

Random Forest Regressor	`rfr = RandomForestRegressor()`
	
	`param_grid = {`
	`'n_estimators': [100, 200, 300],`
	`'max_features': [ 'sqrt', 'log2', None],`
	`'max_depth': [None, 10, 20, 30],`
	`'min_samples_split': [2, 5, 10],`
	`'min_samples_leaf': [1, 2, 4],`
	`'bootstrap': [True, False]`
	`}`
		

Gradient Boosting Regressor	`gb = GradientBoostingRegressor()`
	
	`param_grid = {`
	`'n_estimators': [100, 200, 300],`
	`'learning_rate': [0.001, 0.01, 0.05],`
	`'max_depth': [3, 4, 5, 6],`
	`'min_samples_split': [2, 5, 10],`
	`'min_samples_leaf': [1, 2, 4],`
	`'max_features': [ 'sqrt', 'log2']`
	`}`
		

XGB Regressor	`xg = XGBRegressor(objective='reg:squarederror')`
	
	`param_grid = {`
	`'n_estimators': [100, 200, 300],`
	`'learning_rate': [0.001, 0.0, 0.1],`
	`'max_depth': [3, 4, 5, 6],`
	`'min_child_weight': [1, 3, 5],`
	`'subsample': [0.6, 0.8, 1.0],`
	`'colsample_bytree': [0.6, 0.8, 1.0]`
	`}`
		

AdaBoost Regressor	`ada_boost = AdaBoostRegressor(random_state=42)`
	
	`param_grid = {`
	`'n_estimators': [10 ,50, 100, 200, 300],`
	`'learning_rate': [0.001,0.01, 0.1, 1.0]`
	`}`	

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Linear Regressor	Linear Regression - Root Mean Squared Error: 0.14607218107119907 Linear Regression - R^2 Score: 0.1964180779997876
Decision Tree Regressor	DecisionTreeRegressor - Root Mean Squared Error: 0.12765849855752864 DecisionTreeRegressor - R^2 Score: 0.38624563411216717
Random Forest Regressor	RandomForestRegressor - Root Mean Squared Error: 0.11211841717670257 RandomForestRegressor - R^2 Score: 0.5265773347637333
Gradient Boosting Regressor	GradientBoostingRegressor - Root Mean Squared Error: 0.11485062899977652 GradientBoostingRegressor - R^2 Score: 0.503222536034073
XGB Regressor	XGBRegressor - Root Mean Squared Error: 0.11716056848197681 XGBRegressor - R^2 Score: 0.4830386550037017
AdaBoost Regressor	AdaBoostRegressor - Root Mean Squared Error: 0.12741167428022274 AdaBoostRegressor - R^2 Score: 0.3886166948577655

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest Regressor	The Random Forest Regressor achieved the lowest Root Mean Squared Error (RMSE) among all the models I tested. RMSE measures the differences between the predicted and actual values, so a lower RMSE indicates that the model's predictions are closer to the actual values, suggesting higher accuracy. Additionally, the Random Forest Regressor had the highest R-squared (R^2) score. The R^2 score shows how well the model explains the variance in the target variable. A higher R^2 score means the model is better at capturing the underlying patterns in the data, which is crucial for accurate predictions