Model Optimization and Tuning Phase Template

Date	05 July 2024
Team ID	SWTID1720196555
Project Title	Ecommerce Shipping Prediction Using Machine Learning
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):

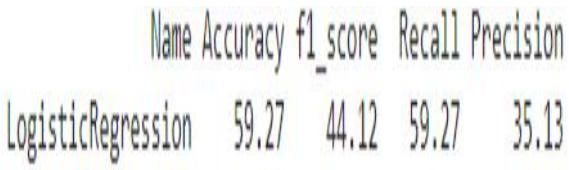
Model	Tuned Hyperparameters	Optimal Values
SVC	{'C': [0.1, 1, 10], 'kernel': ['linear', 'rbf']}	accuracy f1 Recall Precision Support Vector Classifier 65.86 66.12 65.86 66.83
Logistic RegressionCV	{'Cs': [1, 10, 100]}	1 LogisticRegressionCV 63.45 63.61 63.45 63.84

XGBoost	{'n_estimators': [50, 100], 'learning_rate': [0.01, 0.1]}	XGBoost 67.68 67.74 67.68 72.06
Ridge Classifier	{'alpha': [0.1, 1, 10]}	Ridge Classifier 60.95 52.94 60.95 60.20
Knn	{'n_neighbors': [3, 5, 7]}	Knn 64.09 64.32 64.09 64.80
Random Forest	{'n_estimators': [50, 100], 'max_depth': [None, 10, 20]}	Random Forest 68.00 67.94 68.00 73.25
Logistic Regression	{'C': [0.1, 1, 10]}),	Name Accuracy f1_score Recall Precision LogisticRegression 59.27 44.12 59.27 35.13

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
SVC	-- SVM classifier Train Score: 0.5976815547221275 Test Score: 0.5927272727272728 Accuracy: 0.5927272727272728 F1 Score: 0.44116230801162304 Recall: 0.5927272727272728 Precision: 0.3513256198347108	Support Vector Classifier 65.86 66.12 65.86 66.83

Logistic Regression CV	-- Logistic Regression CV Train Score: 0.6329128310035231 Test Score: 0.6345454545454545 Accuracy: 0.6345454545454545 F1 Score: 0.6360729552579082 Recall: 0.6345454545454545 Precision: 0.638414575620458	1 LogisticRegressionCV 63.45 63.61 63.45 63.84
XGBoost	-- XGBoost Train Score: 0.9496533697011024 Test Score: 0.649090909090909 Accuracy: 0.649090909090909 F1 Score: 0.6509629085644186 Recall: 0.649090909090909 Precision: 0.6543703654959536	XGBoost 67.68 67.74 67.68 72.06
Ridge Classifier	-- Ridge Classifier Train Score: 0.5976815547221275 Test Score: 0.5927272727272728 Accuracy: 0.5927272727272728 F1 Score: 0.44116230801162304 Recall: 0.5927272727272728 Precision: 0.3513256198347108	Ridge Classifier 60.95 52.94 60.95 60.20
KNN	-- KNN Train Score: 0.7787248550971702 Test Score: 0.6331818181818182 Accuracy: 0.6331818181818182 F1 Score: 0.6344049935755085 Recall: 0.6331818181818182 Precision: 0.6361072836893792	Knn 64.09 64.32 64.09 64.80
Random Forest	-- Random Forest Train Score: 1.0 Test Score: 0.6668181818181819 Accuracy: 0.6668181818181819 F1 Score: 0.6696708789669983 Recall: 0.6668181818181819 Precision: 0.6812979657035472	Random Forest 68.00 67.94 68.00 73.25

Logistic Regression	-- Logistic Regression Train Score: 0.5976815547221275 Test Score: 0.5927272727272728 Accuracy: 0.5927272727272728 F1 Score: 0.44116230801162304 Recall: 0.5927272727272728 Precision: 0.3513256198347108	

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	The selection of Random Forest as the final model is justified by its high accuracy, robustness, and ability to handle complex data relationships. Its performance during hyperparameter tuning, coupled with its capacity to minimize overfitting and provide feature importance insights, ensures that it meets the project objectives effectively. Therefore, Random Forest is the optimal choice for delivering accurate and reliable eCommerce predictions.