Model Optimization and Tuning Phase Template

Date	15 July 2024
Team ID	SWTID1720151584
Project Title	E-Commerce 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
Random Forest Classifier	'n_estimators':[150,500], 'criterion':['gini', 'entropy'], 'max_depth' : [7], 'max_features' : [60,80,100]	{'criterion': 'gini', 'max_depth': 7, 'max_features': 80, 'n_estimators': 150}
XG Boost Classifier	'min_child_weight' : [10,20], 'gamma' : [1.5,2.0,2.5], 'colsample_bytree' : [0.6,0.8,0.9], 'max_depth' : [4,5,6]	{colsample_bytree': 0.9, 'gamma': 2.5, 'max_depth': 5, 'min_child_weight' : 20}

Logistic Regression	'C' : [6,8,10,15,20], 'max_iter' : [60,80,100]	{'C': 6, 'max_iter': 60}
Support Vector Machine	'kernel' : ['ploy','rbf'], 'C' : [10,13], 'gamma' : [4,5], 'tol' : [1e-1,1e-2,1e-3]	{'C': 10, 'gamma': 4, 'kernel': 'rbf', 'tol': 0.01}

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric

Random Forest Classifier	Accuracy : 0.67
	F1 Score :0.71
	Confusion Matrix:
	`[[ 894  676]`
	`[ 418 1312]]`
		Accuracy : 0.69
		F1 Score :0.67
		Confusion Matrix:
		`[[1222  90]`
		`[ 949 1039]]`

XG Boost Classifier	Accuracy :0.65
	F1 Score :0.59
	Confusion Matrix:
	`[[1297 1149]`
	`[  15  839]]`
		Accuracy : 0.68
		F1 Score :0.67
		Confusion Matrix:
		`[[1182  130]`
		`[ 918 1070]]`

Logistic Regression	Accuracy : 0.64
	F1 Score :0.69
	Confusion Matrix:
	`[[ 766  546]`
	`[ 649 1339]]`
		Accuracy : 0.64
		F1 Score :0.69
		Confusion Matrix:
		`[[ 776  536]`
		`[ 653 1335]]`

Support Vector Machine	Accuracy : 0.66 F1 Score :0.65 Confusion Matrix: [[1175 137] [976 1012]]	Accuracy : 0.64 F1 Score :0.69 Confusion Matrix: [[776 536] [653 1335]]

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest Classifier(Tuned)	A Random Forest classifier is chosen for its high accuracy, robustness against overfitting, versatility in handling both classification and regression tasks, feature importance insights, ability to handle missing data, scalability with large datasets, and strong generalization to unseen data.