

Model Optimization and Tuning Phase Template

Date	06 July 2024
Team ID	SWTID1720260935
Project Title	ECommerce Shipping Prediction Using Machine Learning
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

At this stage, machine learning models are optimized and tuned to achieve optimal performance. To maximize expected accuracy and efficiency, it involves evaluating performance measures, fine-tuning hyperparameters, and supporting the final model selection.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
SVM	c, kernel, gamma	1.0, rbf, 0.01
random forest	n_Estimators, Criterion, Max_Depth, Max_features	none, 1e-9
KNN	n_neighbors, weights, algorithm, p	25, uniform, auto, 2
XGBoost	booster	gbtree

ANN	Units, kernel_initialiser, activation	<p>Input layer: 16, 'random_uniform', 'relu'</p> <p>First Hidden Layer:</p> <p>16, 'random_uniform', 'relu'</p> <p>Second Hidden Layer: 8, 'random_uniform', 'relu'</p> <p>Output layer: 1, 'random_uniform', 'relu'</p>
-----	---------------------------------------	--

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
SVM	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.54 0.87 0.66 1312 1 0.85 0.51 0.64 1988 accuracy 0.65 3300 macro avg 0.70 0.69 0.65 3300 weighted avg 0.73 0.65 0.65 3300 Confusion Matrix: [[1139 173] [977 1011]] </pre> <p>:</p>	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.52 0.50 0.51 1312 1 0.68 0.70 0.69 1988 accuracy 0.62 3300 macro avg 0.60 0.60 0.60 3300 weighted avg 0.62 0.62 0.62 3300 Confusion Matrix: [[655 657] [594 1394]] </pre>

random forest	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.57 0.77 0.65 1312 1 0.80 0.61 0.69 1988 accuracy 0.67 3300 macro avg 0.68 0.69 0.67 3300 weighted avg 0.71 0.67 0.68 3300 Confusion Matrix: [[1009 303] [774 1214]] </pre>	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.56 0.94 0.70 1312 1 0.93 0.51 0.66 1988 accuracy 0.68 3300 macro avg 0.75 0.73 0.68 3300 weighted avg 0.78 0.68 0.68 3300 Confusion Matrix: [[1235 77] [965 1023]] </pre>
KNN	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.53 0.69 0.60 1312 1 0.74 0.59 0.66 1988 accuracy 0.63 3300 macro avg 0.63 0.64 0.63 3300 weighted avg 0.66 0.63 0.63 3300 Confusion Matrix: [[905 407] [812 1176]] </pre>	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.54 0.73 0.62 1312 1 0.77 0.58 0.66 1988 accuracy 0.64 3300 macro avg 0.65 0.66 0.64 3300 weighted avg 0.67 0.64 0.64 3300 Confusion Matrix: [[957 355] [829 1159]] </pre>
XGBoost	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.56 0.70 0.62 1312 1 0.76 0.64 0.70 1988 accuracy 0.66 3300 macro avg 0.66 0.67 0.66 3300 weighted avg 0.68 0.66 0.67 3300 Confusion Matrix: [[916 396] [718 1270]] </pre>	<p>Accuracy,F1 Score:</p> <pre> Classification Report: precision recall f1-score support 0 0.56 0.70 0.62 1312 1 0.76 0.64 0.70 1988 accuracy 0.66 3300 macro avg 0.66 0.67 0.66 3300 weighted avg 0.68 0.66 0.67 3300 Confusion Matrix: [[916 396] [718 1270]] </pre>

ANN	Accuracy,F1 Score:					Accuracy,F1 Score:				
	Classification Report:					Classification Report:				
		precision	recall	f1-score	support		precision	recall	f1-score	support
	0	0.52	0.67	0.59	1312	0	0.52	0.67	0.59	1312
	1	0.73	0.59	0.66	1988	1	0.73	0.59	0.66	1988
	accuracy			0.63	3300	accuracy			0.63	3300
	macro avg	0.63	0.63	0.62	3300	macro avg	0.63	0.63	0.62	3300
	weighted avg	0.65	0.63	0.63	3300	weighted avg	0.65	0.63	0.63	3300
	Confusion Matrix:					Confusion Matrix:				
	[[884 428] [806 1182]]					[[885 427] [806 1182]]				

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	<p>Due to its higher performance metrics, the Random Forest model was selected as the most optimized model. It was successful in generating accurate forecasts, as evidenced by its peak accuracy of 68.42%.</p> <p>It also demonstrated a high precision score of 93.00%, demonstrating its consistency in accurately detecting true positives. The ensemble method used by Random Forest reduces overfitting and enhances generalization to fresh data. Because of these features, Random Forest is the best option for improving delivery time estimates while adhering to the project's goals.</p>