

Model Optimization and Tuning Phase Template

Date	11 July 2024
Team ID	SWTID1720369851
Project Title	Ecommerce Shipping 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
Logistic Regression	penalty, C, solver	penalty = ['l2', 'l1', 'elasticnet'] C = [0.0001, 0.001, 0.002] solver = ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga']
Decision Tree Classifier	max_depth, min_samples_split, min_samples_leaf, max_features, Criterion, splitter	max_depth = [int(x) for x in np.linspace(1, 110, num = 30)]

		min_samples_split = [2, 5, 10, 100] min_samples_leaf = [1, 2, 4, 10, 20, 50] max_features = ['auto', 'sqrt'] criterion = ['gini', 'entropy'] splitter = ['best', 'random']
Random Forest	param_name, param_range, parameter_range, cv	param_name = 'n_estimators', param_range = parameter_range, cv = 5
SVM	NA	NA
XGBoost	NA	NA
KNN	n_neighbors, p, algorithm	n_neighbors = list(range(2,30)) p=[1,2] algorithm = ['auto', 'ball_tree', 'kd_tree', 'brute']

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Logistic Regression	Confusion Matrix : [[531 377] [437 855]] Accuracy_Score: 63.0 % F1 Score: 67.75 Precision Score: 69.399 Recall Score : 66.176 AUC Score : 62.328	Confusion Matrix : [[512 396] [369 923]] Accuracy_Score: 65.227 % F1 Score: 70.701 Precision Score: 69.977 Recall Score : 71.44 AUC Score : 63.914
Decision Tree Classifier	Confusion Matrix : [[508 400] [392 900]] Accuracy_Score: 64.0 % F1 Score: 69.444 Precision Score: 69.231 Recall Score : 69.659 AUC Score : 62.803	Confusion Matrix : [[559 349] [428 864]] Accuracy_Score: 64.682 % F1 Score: 68.982 Precision Score: 71.228 Recall Score : 66.873 AUC Score : 64.218
Random Forest	Confusion Matrix : [[611 297] [446 846]] Accuracy_Score: 66.227 % F1 Score: 69.487 Precision Score: 74.016 Recall Score : 65.48 AUC Score : 66.385	NA

SVM	<p>Confusion Matrix : [[566 342] [402 890]]</p> <p>Accuracy_Score: 66.182 %</p> <p>F1 Score: 70.523 Precision Score: 72.24 Recall Score : 68.885 AUC Score : 65.61</p>	NA
XGBoost	<p>Confusion Matrix : [[567 341] [442 850]]</p> <p>Accuracy_Score: 64.409 %</p> <p>F1 Score: 68.466 Precision Score: 71.369 Recall Score : 65.789 AUC Score : 64.117</p>	NA
KNN	<p>Confusion Matrix : [[556 352] [409 883]]</p> <p>Accuracy_Score: 65.409 %</p> <p>F1 Score: 69.885 Precision Score: 71.498 Recall Score : 68.344 AUC Score : 64.789</p>	<p>Confusion Matrix : [[686 222] [533 759]]</p> <p>Accuracy_Score: 65.682 %</p> <p>F1 Score: 66.784 Precision Score: 77.37 Recall Score : 58.746 AUC Score : 67.148</p>

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
-------------	-----------

Random Forest	<p>Based on the results of each model algorithm, I decided to use the Random Forest model because this model has a fairly balanced combination of AUC and Recall scores and the results are more interpretable.</p> <p>It is delivering high predictive accuracy by aggregating the results of multiple decision trees, reducing the likelihood of overfitting.</p> <p>It is suitable for large datasets due to its parallel processing capabilities. Balancing the bias and variance by averaging multiple decision trees, leading to a more generalized model.</p> <p>These attributes make Random Forest a reliable and effective choice.</p>
---------------	--