Model Optimization and Tuning Phase Template

Date	07 July 2024
Team ID	SWTID1720451040
Project Title	ECommerce Shipping Prediction Using Machine Learning
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

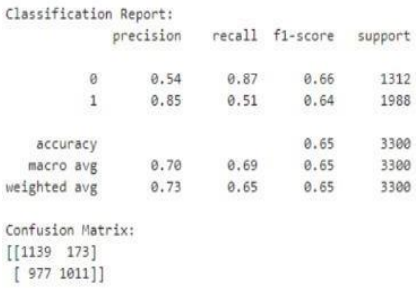
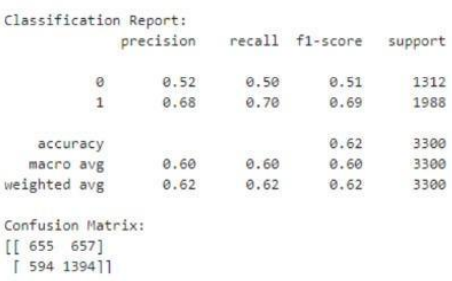
In order to attain optimal performance, machine learning models undergo optimization and tuning at this phase. It entails fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection in order to increase anticipated accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

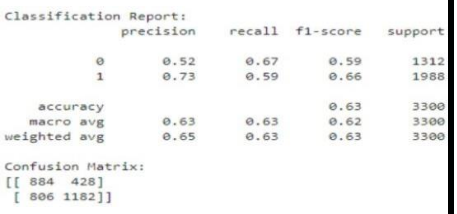
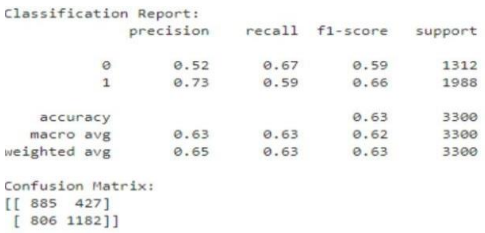
Model	Tuned Hyperparameters	Optimal Values
SVM	c, kernel, gamma	1.0, rbf, 0.01
random forest	n_Estimators, Criterion, Max_Depth, Max_features	none,1e-9
KNN	n_neighbors, weights, algorithm, p	25, uniform, auto, 2
XGBoost	booster	gbtree

ANN	Units, kernel_initialiser, activation	Input layer: 16, ‘random_uniform’, ‘relu’ First Hidden Layer: 16, ‘random_uniform’, ‘relu’ Second Hidden Layer: 8, ‘random_uniform’, ‘relu’ Output layer: 1, ‘random_uniform’, ‘relu’

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
SVM	Accuracy,F1 Score:  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]] :	Accuracy,F1 Score:  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]]

random forest	Accuracy,F1 Score: 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]]	Accuracy,F1 Score: 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]]
KNN	Accuracy,F1 Score: 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]]	Accuracy,F1 Score: 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]]
XGBoost	Accuracy,F1 Score: 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]]	Accuracy,F1 Score: 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]]

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

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	The Random Forest model was chosen as the most optimized model due to its superior performance metrics. Based on its best accuracy of 68.42%, it was effective in producing accurate forecasts . Additionally, it showed a high precision score of 93.00%, proving its dependability in correctly identifying genuine positives. Random Forest's ensemble approach minimizes overfitting and improves generalization to new data. In keeping with the objectives of the project, Random Forest is the ideal choice for enhancing delivery time projections because of these characteristics.