



Model Optimization and Tuning Phase Template

Date	11 July 2024
Team ID	SWTID1720115788
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
Random Forest	## Tuned HyperParameters from sklearn.ensemble import RandomForestClassifier from sklearn.model_selection import GridSearchCV from sklearn.model_selection import GridSearchCV from sklearn.metrics import accuracy_score, classification_report, confusion_matrix rf = RandomForestClassifier(criterion='entropy', random_state=42) param_grid = {'n_estimators': [100, 500, 1000, 2000, 3000, 4000, 5000], 'max_depth': [None, 5, 10, 15, 20], 'min_samples_split': [2, 5, 10, 20, 50], 'class_weight': ['balanced', None]) grid_search = GridSearchCV(rf, param_grid, cv=5, scoring='accuracy') grid_search.fit(X_train, y_train)	Best Random Forest Accuracy: 0.7577272727272727
GRADIENT BOOSTING	<pre>from sklearn.ensemble import GradientBoostingClassifier from sklearn.model_selection import GridSearchCV grid_search = GridSearchCV(GradientBoostingClassifier(random_state=42),{ 'learning_rate': [0.01, 0.1, 1], 'n_estimators': [100, 200, 300], 'max_depth': [3, 5, 7]}, cv=5, scoring='accuracy').fit(X_train, y_train)</pre>	Best Gradient Accuracy: 0.7667272727272727
KNN	## Tuned Hypermeter from sklearn.model_selection import GridSearchCV param_grid = {*n_neighbors': [3, 5, 7, 9, 11],	Best KNN Accuracy: 0.72772727272728





Logistic Regression	<pre>from sklearm.linear_model import LogisticRegression from sklearm.model_selection import GridSearchCV grid_search = GridSearchCV!LogisticRegression(</pre>	Best Logistic Regression Accuracy: 0.6631818181818182
------------------------	--	---

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric					Optimized Metric				
						Best Random For Best Random For				1
	Random Forest Cla		A STATE OF THE PARTY OF THE PAR				recision	recall fi		upport
	pre	cision	recall f	l-score	support	200				83.07
Random		0.00		0.020	1200	0	0.59	0.82	0.69	720
Kandom	0	0.61	0.60	0.61	720	1	0.89	0.73	0.80	1480
	1	0.81	0.81	0.81	1480	accuracy			0.76	2200
Forest	000000000000000000000000000000000000000			0.74	2200	macro avg	0.74	0.77	0.75	2200
01051	accuracy	0.71	0.71	0.74 0.71	2200 2200	weighted avg	0.79	0.76	0.76	2200
	macro avg weighted avg	0.74	0.71	0.71	2200					
						[402 1078]]				
						Best Gradien				
	Gradient Boost	ing Classi	fication (Report:		5-7 93334-0000904-3504-35	Classifica	tion Repor	rt:	support
	3 10 to 10 T	ing Classi		Report:	5upport	Best Gradien		tion Repor		support
	3 10 to 10 T				<i>support</i>	Best Gradien	Classifica	tion Repor	rt:	support 720
GR ADIENT	3 10 to 10 T					Best Gradien Best Gradient	Classifica precision	tion Repor	rt: f1-score	100
GRADIENT	500 ms 3770 co. 15 ms 1 (554 T - 17 to 18 co. 15 d - 17 to 18 co.	precision	recall	f1-score	720	Best Gradient Best Gradient	Classifica precision 0.59	tion Report recall 0.82	rt: f1-score 0.69 0.80	720 1480
	0	orecision 0.59	recall 0.82	f1-score 0.69	720	Best Gradient Best Gradient 2 1 accuracy	Classifica precision 0.59 0.89	recall 0.82 0.73	rt: f1-score 0.69 0.80 0.76	720 1480 2200
	0 1	orecision 0.59	recall 0.82	f1-score 0.69 0.80	720 1480	Best Gradient Best Gradient 2 1 accuracy macro avg	Classifica precision 0.59 0.89	recall 0.82 0.73	rt: f1-score 0.69 0.80 0.76 0.75	720 1480 2200 2200
GRADIENT BOOSTING	0 1 accuracy	0.59 0.89	recall 6.82 6.73	f1-score 0.69 0.80 0.76	720 1480 2200	Best Gradient Best Gradient 2 1 accuracy	Classifica precision 0.59 0.89	recall 0.82 0.73	rt: f1-score 0.69 0.80 0.76	720 1480 2200
	0 1	orecision 0.59	recall 0.82	f1-score 0.69 0.80	720 1480 2200 2200	Best Gradient Best Gradient 2 1 accuracy macro avg	Classifica precision 0.59 0.89 0.74 0.79	tion Repor recall 0.82 0.73 0.77 0.76	rt: f1-score 0.69 0.80 0.76 0.75	720 1480 2200 2200





KNN Classification Report: precision recall f1-score support @ 0.57	0.61 720 0.79 1480								
KNN 0.57 0.59 0.58 720 1 0.80 0.78 0.79 1480 accuracy macro avg 0.68 0.69 0.68 2200 weighted avg 0.72 0.72 0.72 2200 Best KNN Confusion Matrix: [[462 258]									
KNN 1 0.80 0.78 0.79 1480 accuracy accuracy macro avg 0.68 0.69 0.68 2260 weighted avg 0.72 0.72 0.72 2200 Best KNN Confusion Matrix: [[462 258]	0.79 1480								
accuracy accuracy macro avg macro avg weighted avg 0.72 0.72 0.72 0.72 0.72 0.72 0.72 0.72									
accuracy accuracy macro avg 0.68 0.69 0.68 2200 weighted avg 0.72 0.72 2200 Best KNN Confusion Matrix: [[462 258]									
accuracy 0.72 2200 weighted avg 0.74 0.73 macro avg 0.68 0.69 0.68 2200 weighted avg 0.72 0.72 2200 Best KNN Confusion Matrix: [[462 258]	0.73 2200								
macro avg 0.68 0.69 0.68 2200 weighted avg 0.72 0.72 2200 Best KNN Confusion Matrix: [[462 258]	0.70 2200 0.73 2200								
weighted avg 0.72 0.72 2200 Best KNN Confusion Matrix: [[462 258]	0.73 2200								
[[462 258]									
Logistic Regression Classification Report: Best Logistic Regression Accuracy: 0.6631 Best Logistic Regression Classification R									
precision recall f1-score support precision recall f1-sco									
Logistic 0 0.46 0.25 0.32 720 0 0.61 0.60 0.	100775								
Logistic 0 0.46 0.25 0.32 720 1 0.81 0.81 0.	1480								
1 0 70 0 05 0 77 1400	2200								
1 0.70 0.85 0.77 1480									
1 0.70 0.85 0.77 1480 accuracy 0.71 0.71 0.71 0.71 0.71 0.71 0.71 0.71	2200								
1 0.70 0.85 0.77 1480 accuracy 0.									
Regression 1 0.70 0.85 0.77 1480 accuracy macro avg 0.71 0.71 0.71 0.71 0.71 0.71 0.71 0.71									
Regression 1 0.70 0.85 0.77 1480 accuracy 0.71 0.71 0.71 0. macro avg 0.71 0.71 0. weighted avg 0.74 0.74 0.									
Logistic Regression Classification Report: Best Logistic Regression Classification R	ort:								

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
	The Gradient Boosting model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning.
Gradient	Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying
Boosting	its selection as the final model.