Model Development Phase Template

Date	5 July 2024
Team ID	SWTID1720082658
Project Title	E-commerce Shipping Prediction Using Machine Learning
Maximum Marks	6 Marks

Model Selection Report

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Random Forest Classifier	constructs multiple decision trees and combines their predictions to improve accuracy and robustness. It can capture interactions between various features such as carrier reliability, distance, and shipment size.	'classifier_max_depth': 10, 'classifier__min_samples_leaf': 1, 'classifier__min_samples_split': 2, 'classifier__n_estimators': 200	5 5 random forest 65.23 68.76 64.57 73.54
K-Nearest Neighbors Classifier	classifies a shipment's delivery time based on the majority class among its k-nearest neighbors in the feature space. Features might include past delivery times, distances, and shipping conditions.	'classifier_metric': 'euclidean', 'classifier__n_neighbors': 9, 'classifier__p': 1, 'classifier__weights': 'uniform'	4 4 knn 63.41 68.71 67.79 69.66

Model	Description	Hyperparameters	Unnamed: 0		Name	Accuracy	F1_score	Recall	Precision
Logistic Regression	models the probability that a shipment will be delivered on time based on various features such as distance, carrier performance, and historical delivery times.	'classifier__C': 0.01, 'classifier__max_iter': 100, 'classifier__penalty': 'l2', 'classifier__solver': 'liblinear'	0	0	logistic regression	64.05	69.64	69.56	69.72
XGB Classifier	ensemble learning method that builds multiple decision trees sequentially to improve prediction accuracy. It can capture complex relationships and interactions between features such as seasonality, distance, and carrier performance.	'classifier__learning_rate': 0.01, 'classifier__max_depth': 5, 'classifier__n_estimators': 200, 'classifier__subsample': 0.7	2	2	XGBoost	64.64	70.42	71.01	69.83
Support Vector Classifier	finds the optimal hyperplane that separates classes in the feature space, useful for distinguishing between on-time and late deliveries based on features like carrier performance and shipping distance.	'classifier__C': 10, 'classifier__gamma': 'auto', 'classifier__kernel': 'poly'	6	6	support vector classifier	66.36	66.24	55.67	81.76
Ridge Classifier	adds L2 regularization to logistic regression, which can help handle multicollinearity among features such as various delivery time factors and shipping routes.	-	3	3	ridge classifier	65.23	68.76	64.57	73.54
Logistic RegressionCV	extends logistic regression by incorporating cross-validation to automatically tune hyperparameters like the regularization strength. This can improve prediction accuracy by preventing overfitting.	-	1	1	logistic regression CV	63.77	70.27	72.24	68.41