Model Development Phase Template

Date	04 June 2024
Team ID	SWTID1720096620
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	This ensemble method combines multiple decision trees to improve accuracy and control over-fitting by averaging their predictions.	'classifier_max_depth': 10, 'classifier__min_samples_leaf': 1, 'classifier__min_samples_split': 2, 'classifier__n_estimators': 200	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.58 0.95 0.72 1379 1 0.93 0.50 0.65 1921 accuracy 0.69 3300 macro avg 0.75 0.73 0.69 3300 weighted avg 0.78 0.69 0.68 3300
K-Nearest Neighbors Classifier	A simple, non-parametric method that classifies a data point based on the majority label among its k-nearest neighbors in the feature space.	'classifier_metric': 'euclidean', 'classifier__n_neighbors': 9, 'classifier__p': 1, 'classifier__weights': 'uniform'	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.57 0.81 0.67 1379 1 0.80 0.56 0.66 1921 accuracy 0.66 3300 macro avg 0.68 0.68 0.66 3300 weighted avg 0.70 0.66 0.66 3300

Logistic Regression	A linear model used for binary classification that estimates the probability of a binary response based on one or more predictor variables using a logistic function.	'classifier__C': 0.01, 'classifier__max_iter': 100, 'classifier__penalty': 'l2', 'classifier__solver': 'liblinear'	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.55 0.76 0.64 1379 1 0.77 0.56 0.65 1921 accuracy 0.64 3300 macro avg 0.66 0.66 0.64 3300 weighted avg 0.68 0.64 0.64 3300
XGB Classifier	An optimized gradient boosting library designed for speed and performance, which builds an ensemble of decision trees by sequentially minimizing a loss function.	'classifier__learning_rate': 0.01, 'classifier__max_depth': 5, 'classifier__n_estimators': 200, 'classifier__subsample': 0.7	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.58 0.96 0.72 1379 1 0.94 0.50 0.65 1921 accuracy 0.69 3300 macro avg 0.76 0.73 0.69 3300 weighted avg 0.79 0.69 0.68 3300
Support Vector Classifier	A classifier that constructs a hyperplane in a high-dimensional space to separate different classes with maximum margin, often using kernel functions for non-linear separation.	'classifier__C': 10, 'classifier__gamma': 'auto', 'classifier__kernel': 'poly'	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.56 0.92 0.70 1379 1 0.90 0.49 0.63 1921 accuracy 0.67 3300 macro avg 0.73 0.71 0.67 3300 weighted avg 0.76 0.67 0.66 3300
Decision Tree Classifier	A model that splits the data into subsets based on feature values, creating a tree structure where each leaf represents a class label and each node represents a decision rule.	'classifier__criterion': 'gini', 'classifier__max_depth': 10, 'classifier__min_samples_leaf': 4, 'classifier__min_samples_split': 2	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.57 0.89 0.70 1379 1 0.87 0.53 0.66 1921 accuracy 0.68 3300 macro avg 0.72 0.71 0.68 3300 weighted avg 0.74 0.68 0.67 3300
Naive Bayes Classifier	A probabilistic classifier based on Bayes' theorem, which assumes independence among features and calculates the probability of each class given the input features.	'classifier__var_smoothing': 1e-09	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.55 0.89 0.68 1379 1 0.86 0.48 0.62 1921 accuracy 0.65 3300 macro avg 0.70 0.68 0.65 3300 weighted avg 0.73 0.65 0.64 3300

AdaBoost Classifier	An ensemble method that combines multiple weak classifiers, typically decision trees, by weighting them according to their accuracy and iteratively improving the model.	'classifier__learning_rate': 1, 'classifier__n_estimators' : 200	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.58 0.89 0.71 1379 1 0.87 0.54 0.67 1921 accuracy 0.69 3300 macro avg 0.73 0.72 0.69 3300 weighted avg 0.75 0.69 0.68 3300
Gradient Boost Classifier	An ensemble technique that builds models sequentially, each new model correcting the errors of the previous ones, and combines them to make a final prediction.	'classifier__learning_rate': 0.01, 'classifier__max_depth': 5, 'classifier__n_estimators': 200, 'classifier__subsample': 0.9	Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.58 0.95 0.72 1379 1 0.93 0.51 0.66 1921 accuracy 0.69 3300 macro avg 0.76 0.73 0.69 3300 weighted avg 0.79 0.69 0.68 3300