Model Development Phase Template

Date	12 March 2024
Team ID	SWTID1720089323
Project Title	Ecommerce 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
Logistic Regression	Logistic regression is a type of linear regression used for binary classification problems, where the outcome (dependent variable) is binary (0 or 1). It models the probability of the default class (usually 1) using the logistic function.	-	Accuracy=62.94 F1_score= 66.37 Recall=60.71 Precision=73.20
Logistic Regression CV	LogisticRegressionCV is a logistic regression model with built-in cross-validation to automatically determine the best regularization parameter C from a predefined list (Cs). This helps in finding a balance between underfitting and overfitting by selecting the optimal regularization strength.	-	Accuracy=62.61 F1_score=65.70 Recall=59.46 Precision=73.42

XGBoost Classifier	XGBoost (Extreme Gradient Boosting) is an efficient and scalable implementation of gradient boosting framework. It's widely used for classification and regression tasks and often provides better performance compared to other algorithms.	-	Accuracy=66.24 F1_score=69.51 Recall=63.88 Precision=76.23
Ridge Classifier	Ridge Classifier is a linear classifier that uses Ridge regression to handle multicollinearity in the data and improve model generalization.	-	Accuracy=62.82 F1_score=66.15 Recall=60.31 Precision=73.24
K-Nearest Neighbors Classifier	K-Nearest Neighbors is a simple and effective classification algorithm that stores all available cases and classifies new cases based on a similarity measure	-	Accuracy=63.06 F1_score=65.86 Recall=59.15 Precision=74.29
Random Forest Classifier	Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.	-	Accuracy=67.09 F1_score=68.95 Recall=60.66 Precision=79.87
Support Vector Machine Classifier	Support Vector Classifier (SVC) is a powerful supervised learning algorithm used for classification tasks. It works by finding the hyperplane that best separates the classes in the feature space.	-	Accuracy=65.15 F1_score=63.75 Recall=50.86 Precision=85.39