Model Development Phase Template

Date	1 July 2024
Team ID	SWTID1720434734
Project Title	Ecommerce Shipping Prediction
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)
Logistic Regression	Logistic Regression is a linear model commonly used for binary classification tasks. It predicts the probability of an instance belonging to a particular class using a logistic function.	-	Accuracy score = 59%
Logistic Regression CV	Logistic Regression with Cross-Validation (CV) is an extension of Logistic Regression that uses cross-validation to find the best hyperparameters, typically the regularization strength.	-	Accuracy score = 62%

XGBoost	XGBoost (Extreme Gradient Boosting) is an advanced implementation of gradient boosting designed for speed and performance. It builds an ensemble of trees sequentially to minimize the classification error.	-	Accuracy score = 67%
Ridge Classifier	Ridge Classifier applies L2 regularization to the Logistic Regression model, which penalizes large coefficients and helps prevent overfitting.	-	Accuracy score = 59%
KNN	KNN is a non-parametric, instance-based learning algorithm that classifies instances based on the majority class among the k-nearest neighbors in the feature space.	-	Accuracy score = 63%
Random Forest	Random Forest is an ensemble method that constructs multiple decision trees during training and outputs the mode of their predictions for classification tasks. It reduces overfitting by averaging multiple trees.	-	Accuracy score = 67%
SVM Classifier	SVM is a supervised learning algorithm that finds the optimal hyperplane to separate classes in the feature space. It can use different kernel functions to handle non-linear relationships.	-	Accuracy score = 59%