Model Development Phase Template

Date	**11 July 2024**
Team ID	**SWTID1720163281**
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 Metric (e.g., Accuracy, F1 Score)
Logistic Regression	Logistic Regression is a statistical model predicting the probability that a certain event occurs, which would be yeah or no, based on input variables through the use of a logisticfunction mapping the input to a probability between 0 and 1. It is one of the most widely used and effective algorithms for binary classification problems, such as classifying whether a customer will churn or not, or whether a tumor is malignant or benign.	test_size=0.2	Accuracy Score = 65.681

Random Forest	One of the important algorithms of ensemble learning is a Random Forest, which combines many decision trees aimed at enhancing the accuracy and robustness of predictions by reducing overfitting and increasing model generalization capacity. In this way, Random Forest achieves high accuracy beyond that by individual decision trees, hence its popularity for classification and regression tasks.	n_estimators=100,random_state=42,criterion='entropy'	Accuracy Score = 74.318
GRADIENT BOOSTING	Gradient Boosting is an ensemble learning algorithm that combines many weak models to get a strong predictive model by iteratively training each model based on the error derived in the previous model, hence improving in accuracy and robustness. On the other end, Gradient Boosting can achieve state-of-the-art performance for a large group of machine learning tasks by utilizing multiple models, mostly with large datasets and complex interactions.	n_estimators=100,learning_rate=0.1,random_state=42,	Accuracy Score = 75.772
KNN	K-Nearest Neighbors (KNN) is a simple yet effective machine learning algorithm that predicts the output of a new instance by finding the most similar instances in the training data and voting on their labels, resulting in accurate	n_neighbors=5,weights=uniform,leaf_size=30	Accuracy Score = 72.000

	predictions for classification and regression tasks.		