

Model Development Phase Template

Date	12 July 2024
Team ID	SWTID1720108776
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	It is a statistical method primarily used for binary classification tasks. The model uses a logistic function (also known as the sigmoid function) to map the linear combination of features to a probability score between 0 and 1.	–	Accuracy Score = 67%
Random Forest	It is an ensemble learning technique based on decision trees. It combines multiple decision trees to improve overall accuracy and reduce	Criterion= 'entropy'	Accuracy Score = 73%

	overfitting. Each tree in the forest is trained on random subset of features. The final prediction is obtained by aggregating the predictions of individual trees. Random Forest handles non-linear relationships well and is robust against noisy data.		
Decision Tree	A hierarchical model. It resembles a flowchart, with nodes representing decision stages. Internal nodes correspond to attribute tests, and leaf nodes indicate class labels. Decision trees are interpretable and widely used for classification and regression tasks.	Criterion= 'entropy'	Accuracy Score = 70%
K-Nearest Neighbours	k-NN is an instance-based learning algorithm. Given a new data point, it identifies the k nearest neighbors (based on a distance metric like Euclidean distance) and assigns the majority class label among those neighbors.	–	Accuracy Score = 70%
SVM(Support Vector Machine)	SVM creates a hyperplane to separate data into classes. It maximizes the margin between these classes, aiming to find the best separation boundary. It works well for both linearly separable and non-linearly separable data.	–	Accuracy Score = 73%

XG Boost	XGBoost is a powerful gradient boosting algorithm that builds an ensemble of weak learners, typically decision trees. During training, it optimizes a loss function by iteratively adding trees to minimize the error.	–	Accuracy Score = 72%
----------	--	---	----------------------