

Model Development Phase Template

Date	17 July 2024
Team ID	SWTID1720190389
Project Title	E-Commerce 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
Logistic Regression	Logistic Regression is a statistical method used for binary classification problems. It models the probability that a given input belongs to a particular class by applying a logistic function (sigmoid) to a linear combination of input features. Despite its name, it is used for classification rather than regression tasks.	<pre>param_grids = { 'LogisticRegression': { 'solver': ['liblinear', 'lbfgs'], 'C': [0.01, 0.1, 1, 10, 100] }, }</pre>
Gaussian NB	Gaussian Naive Bayes is a variant of the Naive Bayes classifier that assumes the features follow a normal (Gaussian) distribution. It calculates the probability of each class given the input features and selects the class with the highest probability. It is particularly effective for problems where the feature distributions are well-approximated by a Gaussian distribution.	<pre>param_grids = { 'GaussianNB': { 'var_smoothing': [1e-09, 1e-08, 1e-07, 1e-06, 1e-05] }, }</pre>
KNN	Classifies based on nearest neighbors; adapts well to data patterns, effective for local variations in loan approval criteria.	<pre>param_grids = {</pre>

		<pre>'KNeighborsClassifier': { 'n_neighbors': [3, 5, 7, 9, 11], 'weights': ['uniform', 'distance'] },</pre>
Decision Tree	Simple tree structure; interpretable, captures non-linear relationships, suitable for initial insights into loan approval patterns.	<pre>param_grids = { 'DecisionTreeClassifier': { 'max_depth': [3, 5, 7, 9, None], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4] },</pre>
Random Forest	Ensemble of decision trees; robust, handles complex relationships, reduces overfitting, and provides feature importance for loan approval prediction.	<pre>param_grids = { 'RandomForestClassifier': { 'n_estimators': [50, 100, 200], 'max_depth': [3, 5, 7, 9, None], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4] },</pre>
Gradient Boosting	Gradient boosting with trees; optimizes predictive performance, handles complex relationships, and is suitable for accurate loan approval predictions.	<pre>param_grids = { 'GradientBoostingClassifier': { 'n_estimators': [50, 100, 200], 'learning_rate': [0.01, 0.1, 0.05], 'max_depth': [3, 5, 7, 9] },</pre>
XGBoost	XGBoost (Extreme Gradient Boosting) is an optimized and scalable implementation of gradient boosting for decision trees. It uses a series of weak learners (typically decision trees) to create a strong learner by iteratively minimizing the loss function. XGBoost is known for its speed, performance, and ability to handle large-scale datasets.	<pre>param_grids = { 'XGBClassifier': { 'n_estimators': [50, 100, 200], 'learning_rate': [0.01, 0.1, 0.05], 'max_depth': [3, 5, 7, 9] },</pre>
CatBoost	CatBoost (Categorical Boosting) is a gradient boosting algorithm that is specifically designed to handle categorical features efficiently. It uses ordered boosting to prevent target leakage and applies various techniques to handle categorical	<pre>param_grids = {</pre>

	data without extensive preprocessing. CatBoost is known for its high accuracy and performance on datasets with categorical features.	<pre>'CatBoostClassifier': { 'iterations': [50, 100, 200], 'learning_rate': [0.01, 0.1, 0.05], 'depth': [3, 5, 7, 9] }</pre>
--	--	--

Performance Metric:

Model Evaluation before Hyper Parameter Tuning

	Name	Accuracy	f1_score	Recall	Precision
0	logistic regression	67.62	71.20	68.32	74.32
1	GaussianNB	63.85	55.36	38.28	100.00
2	KNeighborsClassifier	67.03	71.24	69.72	72.83
3	DecisionTreeClassifier	64.71	70.72	72.75	68.80
4	RandomForestClassifier	67.44	70.61	66.77	74.91
5	GradientBoostingClassifier	69.99	68.96	56.91	87.47
6	XGBClassifier	66.89	70.33	67.00	74.01
7	CatBoostClassifier	68.26	70.09	63.51	78.20