

Model Optimization and Tuning Phase Template

Date	15 July 2024
Team ID	SWTID1720086535
Project Title	Ecommerce Shipping Prediction Using Machine Learning
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
RandomFore stClassifier	<pre>#HyperParameter Optimisation for Random Forest rf = RandomForestClassifier() rf_param_grid = { 'n_estimators': [200,300,500], 'criterion': ['entropy', 'gini'], 'max_depth': [7,8,60,80,100], 'max_features': ['sqrt', 'log2'] } rf_cv= GridSearchCV(rf,rf_param_grid, cv=7, scoring="accuracy", n_jobs=-1, verbose=3) rf_cv.fit(xnorm_train,y_train) print("Best Score:" + str(rf_cv.best_score_))</pre>	--
SVM	<pre>#HyperParameter Optimisation for SVM Svc = svm.SVC(random_state=1234) params = { 'kernel': ['poly', 'rbf'], 'C': [10, 13], 'gamma': [4,5], 'tol':[1e-1,1e-2,1e-3] } fitmodel = GridSearchCV(svc, param_grid=params, cv=5, refit=True, scoring="accuracy", n_jobs=-1, verbose=3) fitmodel.fit(xnorm_train, y_train) print(fitmodel.best_estimator_, fitmodel.best_params_, fitmodel.best_score_)</pre>	<p>Fitting 5 folds for each of 24 candidates, totalling 120 fits</p> <p>SVC(C=6, gamma=2, random_state=1234) {'C': 6, 'gamma': 2, 'kernel': 'rbf'} 0.6659045470650132</p>

<p>XGBoost</p>	<pre>#HyperParameter Optimisation for XGBoost params = { 'min_child_weight': [10,20], 'gamma': [1.5, 2.0, 2.5], 'colsample_bytree': [0.6, 0.8, 0.9], 'max_depth': [4,5,6] } xgb = XGBClassifier(learning_rate=0.5, n_estimators=100, objective='binary:logistic', nthread=3) fitmodel = GridSearchCV(xgb, param_grid=params, cv=5, refit=True, scoring="accuracy", n_jobs=-1, verbose=3) fitmodel.fit(xnorm_train, y_train) print(fitmodel.best_estimator_, fitmodel.best_params_, fitmodel.best_score_)</pre>	<p>Fitting 5 folds for each of 84 candidates, totalling 210 fits</p> <pre>XGBClassifier(base_score=None, booster=None, call_hooks=None, colsample_bytree=None, colsample_bynode=None, colsample_byrow=None, device=None, early_stopping_rounds=None, enable_categorical=False, eval_metric=None, feature_types=None, gamma=None, grow_policy=None, importance_type=None, interaction_constraints=None, learning_rate=0.5, max_bin=None, max_cat_threshold=None, max_cat_to_onehot=None, max_delta_step=None, max_depth=5, max_leaves=None, min_child_weight=30, missing nan, monotone_constraints=None, multi_strategy=None, n_estimators=100, n_jobs=None, nthread=3, num_parallel_tree=None, ...) ('colsample_bytree': 0.6, 'gamma': 2.0, 'max_depth': 5, 'min_child_weight': 20) 0.672657653444624</pre>
<p>Logistic Regression CV</p>	<pre>#HyperParameter Optimisation for Logistic Regression lg = LogisticRegressionCV(n_jobs=-1, random_state= 1234) lg_param_grid = { 'cs': [6,8,10,15,20], 'max_iter': [60,80,100] } lg_cv= GridSearchCV(lg,lg_param_grid,cv=5, scoring="accuracy", n_jobs=-1, verbose=3) lg_cv.fit(xnorm_train,y_train)</pre>	<p>Fitting 5 folds for each of 15 candidates, totalling 75 fits</p> <pre>GridSearchCV estimator: LogisticRegressionCV LogisticRegressionCV LogisticRegressionCV(n_jobs=-1, random_state=1234)</pre> <p>Optimal parameters:({'cs': 8, 'max_iter': 60}) Accuracy on test set:0.6359090909090909</p>

Performance Metrics Comparison Report (2 Marks):

	Name	Accuracy	F1_score	Recall	Precision
0	logistic regression	64.05	69.64	69.56	69.72
1	logistic regression CV	63.77	70.27	72.24	68.41
2	XGBoost	64.64	70.42	71.01	69.83
3	ridge classifier	65.23	68.76	64.57	73.54
4	knn	63.41	68.71	67.79	69.66
5	random forest	65.23	68.76	64.57	73.54

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
RandomForestClassifier	The RandomForestClassifier was chosen as the final model due to its superior accuracy after hyperparameter tuning, achieving an optimized accuracy of 0.66 compared to the baseline accuracy of 0.64. The model also demonstrates robustness and generalization capabilities suitable for the project's requirements.