Model Optimization and Tuning Phase Template

Date	3 October 2024
Team ID	LTVIP2024TMID24838
Project Title	Detection of Phishing Websites from URLs
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.

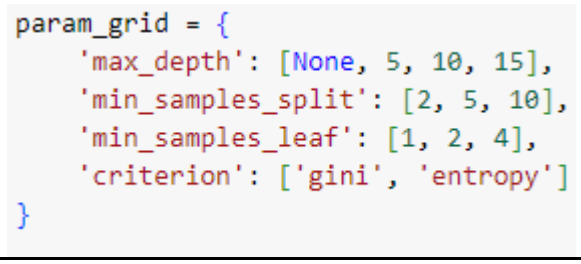
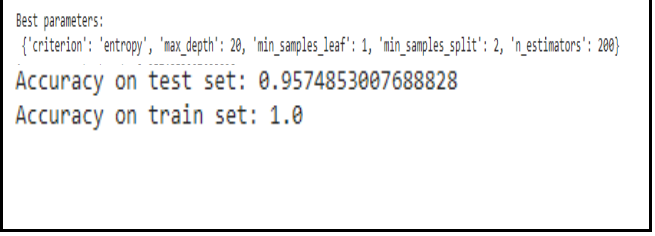
Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values

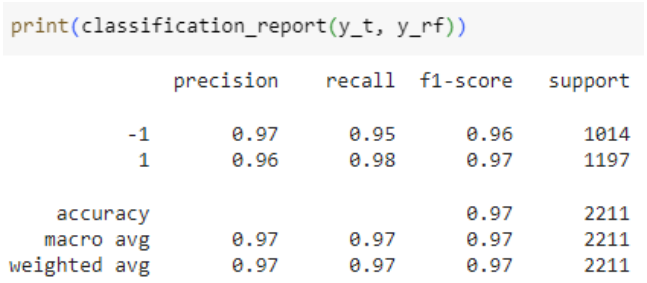
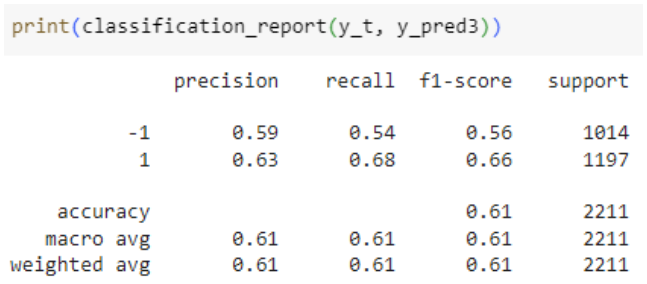
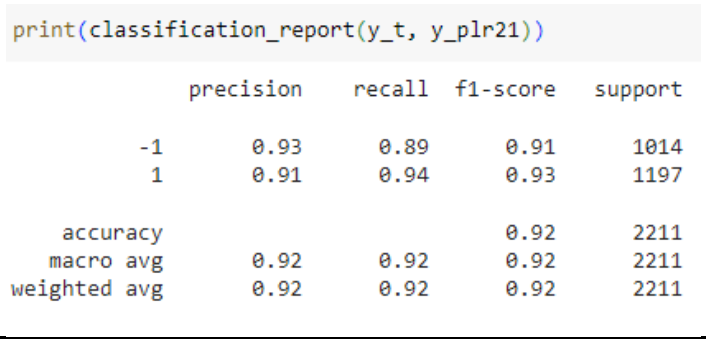
Random Forest	`param_grid = { 'n_estimators': [50, 100, 200], 'max_depth': [None, 10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4], 'criterion':['gini','entropy'] }`	`RandomForestClassifier(criterion='entropy', max_depth=20, n_estimators=200, random_state=0)`
		`test accuracy: 0.9665309814563546`
		`train accuracy 0.9987562189054726`

KNN	`param_grid = { 'n_neighbors': [3, 5, 7, 9, 11], 'weights': ['uniform', 'distance'], 'metric': ['euclidean', 'manhattan'] }`	`KNeighborsClassifier(metric='manhattan', n_neighbors=3, weights='distance')`
		`test accuracy 0.6142017186793306`
		`train accuracy 0.7800768882858435`

Logistic Regression	`param_grid = { 'C': [0.001, 0.01, 0.1, 1, 10, 100], 'penalty': ['l1', 'l2'], 'solver': ['liblinear', 'saga'] }`	`Best parameters: {'C': 1, 'penalty': 'l1', 'solver': 'liblinear'}`
		`Accuracy on test set: 0.9185888738127544`
		`Accuracy on training set: 0.9320443238353686`

Decision Tree Classifier		

Performance Metrics Comparison Report (2 Marks):

Model	**Baseline Metric** **Optimized Metric**
Random Forest	
KNN	
Logistic Regression	

Decision Tree Classifier	`print(classification_report(y_t, y_pred_t))` precision recall f1-score support -1 0.95 0.95 0.95 1014 1 0.96 0.96 0.96 1197 accuracy 0.95 2211 macro avg 0.95 0.95 0.95 2211 weighted avg 0.95 0.95 0.95 2211

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Logistic Regression	▪ According to the above data the model **knn** has the least accuracy. ▪ The **Decision Tree** and **Random Forest** models both has training accuracy of (1.0) which is overfitting of the data, even after the hyperparameter tuning of the models the models show overfitting. ▪ The most suitable models seem to be the **Logistic Regression** among the above models. ▪ The **Logistic Regression** model which has 92% accuracy score.