

Model Optimization and Tuning Phase Template

Date	July 5, 2024
Team ID	739683
Project Title	Customer segmentation 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
Decision tree	<pre># Define the Decision Tree classifier dt_classifier = DecisionTreeClassifier() # Define the hyperparameters and their possible values for tuning param_grid = { 'criterion': ['gini', 'entropy'], 'splitter': ['best', 'random'], 'max_depth': [None, 10, 20, 30, 40, 50], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4] }</pre>	<pre># Evaluate the performance of the tuned model accuracy = accuracy_score(y_test, y_pred) print(f"Optimal Hyperparameters: {best_params}") print(f"Accuracy on Test Set: {accuracy}") Optimal Hyperparameters: {'criterion': 'gini', 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'best'} Accuracy on Test Set: 0.7559333360467</pre>
Random forest	<pre># Define the Random Forest classifier rf_classifier = RandomForestClassifier() # Define the hyperparameters and their possible values for tuning param_grid = { 'n_estimators': [50, 100, 200], 'criterion': ['gini', 'entropy'], 'max_depth': [None, 10, 20, 30], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4], }</pre>	<pre># Evaluate the performance of the tuned model accuracy = accuracy_score(y_test, y_pred) print(f"Optimal Hyperparameters: {best_params}") print(f"Accuracy on Test Set: {accuracy}") Optimal Hyperparameters: {'criterion': 'entropy', 'max_depth': 30, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200} Accuracy on Test Set: 0.751475109540028</pre>

Gradient Boosting	<pre># Define the Gradient Boosting classifier gb_classifier = GradientBoostingClassifier() # Define the hyperparameters and their possible values for tuning param_grid = { 'n_estimators': [50, 100, 200], 'learning_rate': [0.01, 0.1, 0.2], 'max_depth': [3, 4, 5], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4], 'subsample': [0.8, 1.0] }</pre>	<pre># Evaluate the performance of the tuned model y = accuracy_score(y_test, y_pred) # Optimal hyperparameters (best_params) # Accuracy on Test Set: (accuracy)</pre> <pre>Hyperparameters: {'learning_rate': 0.1, 'max_depth': 5, 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 100, 'subsample': 0.8} Accuracy on Test Set: 0.7500000000000001</pre>
-------------------	--	---

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric																														
Decision tree	<pre>print(classification_report(y_test,y_pred))</pre> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>Loan will be Approved</td><td>0.67</td><td>0.68</td><td>0.68</td><td>75</td></tr><tr><td>Loan will not be Approved</td><td>0.74</td><td>0.73</td><td>0.74</td><td>94</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.71</td><td>169</td></tr><tr><td>macro avg</td><td>0.71</td><td>0.71</td><td>0.71</td><td>169</td></tr><tr><td>weighted avg</td><td>0.71</td><td>0.71</td><td>0.71</td><td>169</td></tr></tbody></table> <pre>confusion_matrix(y_test,y_pred)</pre> <pre>array([[51, 24], [25, 69]])</pre>		precision	recall	f1-score	support	Loan will be Approved	0.67	0.68	0.68	75	Loan will not be Approved	0.74	0.73	0.74	94	accuracy			0.71	169	macro avg	0.71	0.71	0.71	169	weighted avg	0.71	0.71	0.71	169
	precision	recall	f1-score	support																											
Loan will be Approved	0.67	0.68	0.68	75																											
Loan will not be Approved	0.74	0.73	0.74	94																											
accuracy			0.71	169																											
macro avg	0.71	0.71	0.71	169																											
weighted avg	0.71	0.71	0.71	169																											
Random forest	<pre>print(classification_report(y_test,y_pred))</pre> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>Loan will be Approved</td><td>0.71</td><td>0.83</td><td>0.77</td><td>75</td></tr><tr><td>Loan will not be Approved</td><td>0.84</td><td>0.73</td><td>0.78</td><td>94</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.78</td><td>169</td></tr><tr><td>macro avg</td><td>0.78</td><td>0.78</td><td>0.77</td><td>169</td></tr><tr><td>weighted avg</td><td>0.78</td><td>0.78</td><td>0.78</td><td>169</td></tr></tbody></table> <pre>confusion_matrix(y_test,y_pred)</pre> <pre>array([[62, 13], [25, 69]])</pre>		precision	recall	f1-score	support	Loan will be Approved	0.71	0.83	0.77	75	Loan will not be Approved	0.84	0.73	0.78	94	accuracy			0.78	169	macro avg	0.78	0.78	0.77	169	weighted avg	0.78	0.78	0.78	169
	precision	recall	f1-score	support																											
Loan will be Approved	0.71	0.83	0.77	75																											
Loan will not be Approved	0.84	0.73	0.78	94																											
accuracy			0.78	169																											
macro avg	0.78	0.78	0.77	169																											
weighted avg	0.78	0.78	0.78	169																											
Gradient boosting	<pre>print(classification_report(y_test,y_pred))</pre> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>Loan will be Approved</td><td>0.73</td><td>0.85</td><td>0.79</td><td>75</td></tr><tr><td>Loan will not be Approved</td><td>0.86</td><td>0.74</td><td>0.80</td><td>94</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.79</td><td>169</td></tr><tr><td>macro avg</td><td>0.80</td><td>0.80</td><td>0.79</td><td>169</td></tr><tr><td>weighted avg</td><td>0.80</td><td>0.79</td><td>0.79</td><td>169</td></tr></tbody></table> <pre>confusion_matrix(y_test,y_pred)</pre> <pre>array([[64, 11], [24, 70]])</pre>		precision	recall	f1-score	support	Loan will be Approved	0.73	0.85	0.79	75	Loan will not be Approved	0.86	0.74	0.80	94	accuracy			0.79	169	macro avg	0.80	0.80	0.79	169	weighted avg	0.80	0.79	0.79	169
	precision	recall	f1-score	support																											
Loan will be Approved	0.73	0.85	0.79	75																											
Loan will not be Approved	0.86	0.74	0.80	94																											
accuracy			0.79	169																											
macro avg	0.80	0.80	0.79	169																											
weighted avg	0.80	0.79	0.79	169																											

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Gradient boosting	The gradient Boosting model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model