

Model Optimization and Tuning Phase Template

Date	12 JULY 2024
Team ID	SWTID1720067156
Project Title	Lymphography Classification Tool
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>param_grid = {'max_depth': [None, 10, 20, 30]} dt = DecisionTreeClassifier(random_state=42) grid_search = GridSearchCV(dt, param_grid, cv=5) grid_search.fit(X_train, y_train) best_dt = grid_search.best_estimator_ y_pred = best_dt.predict(X_test)</pre>	max_depth=10, criterion='gini'
Random Forest	<pre>param_grid = {'n_estimators': [50, 100, 200], 'max_depth': [None, 10, 20, 30]} rf = RandomForestClassifier(random_state=42) grid_search = GridSearchCV(rf, param_grid, cv=5) grid_search.fit(X_train, y_train) best_rf = grid_search.best_estimator_ y_pred = best_rf.predict(X_test)</pre>	n_estimators=100, max_depth=20
KNN	<pre>param_grid = {'n_neighbors': [3, 5, 7, 9, 11]} knn = KNeighborsClassifier() grid_search = GridSearchCV(knn, param_grid, cv=5) grid_search.fit(X_train, y_train) best_knn = grid_search.best_estimator_ y_pred = best_knn.predict(X_test)</pre>	n_neighbors=5, metric='minkowski'
Gradient Boosting	<pre>param_grid = {'n_estimators': [50, 100, 200], 'learning_rate': [0.01, 0.1, 0.2]} gb = GradientBoostingClassifier(random_state=42) grid_search = GridSearchCV(gb, param_grid, cv=5) grid_search.fit(X_train, y_train) best_gb = grid_search.best_estimator_ y_pred = best_gb.predict(X_test)</pre>	learning_rate=0.1, n_estimators=100

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric	Confusion Matrix
Decision Tree	Decision Tree Accuracy: 0.73 Decision Tree Classification Report: <pre> precision recall f1-score support 1 0.65 0.93 0.76 14 2 0.90 0.64 0.75 14 3 0.00 0.00 0.00 2 accuracy 0.73 30 macro avg 0.52 30 weighted avg 0.72 30 </pre>	Confusion Matrix: <pre> [[9 4 1] [2 11 1] [0 0 3]] </pre>
Random Forest	Random Forest Accuracy: 0.83 Random Forest Classification Report: <pre> precision recall f1-score support 1 0.87 0.93 0.90 14 2 0.80 0.86 0.83 14 3 0.00 0.00 0.00 2 accuracy 0.83 30 macro avg 0.56 30 weighted avg 0.78 30 </pre>	Confusion Matrix: <pre> [[10 3 1] [1 11 3] [0 0 2]] </pre>
KNN	KNN Accuracy: 0.87 KNN Classification Report: <pre> precision recall f1-score support 1 0.88 1.00 0.93 14 2 0.86 0.86 0.86 14 3 0.00 0.00 0.00 2 accuracy 0.87 30 macro avg 0.58 30 weighted avg 0.81 30 </pre>	Confusion Matrix: <pre> [[12 2 0] [2 12 0] [0 0 2]] </pre>
Gradient Boosting	Gradient Boosting Accuracy: 0.87 Gradient Boosting Classification Report: <pre> precision recall f1-score support 1 0.81 0.93 0.87 14 2 0.92 0.79 0.85 14 3 1.00 1.00 1.00 2 accuracy 0.87 30 macro avg 0.91 30 weighted avg 0.87 30 </pre>	Confusion Matrix: <pre> [[11 2 1] [3 10 1] [0 0 3]] </pre>

Final Model Selection Justification (2 Marks):

Model	Reasoning
KNN	Reasoning: The KNN model was selected for its superior accuracy of 87%, indicating its effectiveness in handling local variations in lymphography criteria and providing reliable classification results.