Model Optimization and Tuning Phase Report

Team ID: SWTID1720067156

July 11, 2024

Model Optimization and Tuning Phase

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	`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)`
		max_depth=10, criterion='gini'

Random Forest	`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)`
		n_estimators=100, max_depth=20

KNN	`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)`
		n_neighbors=5, metric='minkowski'

Gradient Boosting	`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)`
		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: 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 0.52 0.50 30 weighted avg 0.72 0.73 0.71 30	Confusion Matrix: [[9 4 1] [2 11 1] [0 0 3]]
Random Forest	Random Forest Accuracy: 0.83 Random Forest Classification Report: 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 0.60 0.57 30 weighted avg 0.78 0.83 0.80 30	Confusion Matrix: [[10 3 1] [1 11 3] [0 0 2]]
KNN	KNN Accuracy: 0.87 KNN Classification Report: 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 0.62 0.60 30 weighted avg 0.81 0.87 0.84 30	Confusion Matrix: [[12 2 0] [2 12 0] [0 0 2]]
Gradient Boosting	Gradient Boosting Accuracy: 0.87 Gradient Boosting Classification Report: 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 0.90 0.90 30 weighted avg 0.87 0.87 0.87 30	Confusion Matrix: [[11 2 1] [3 10 1] [0 0 3]]

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.