Model Optimization and Tuning Phase Report

Date	23 September 2024
Team ID	LTVIP2024TMID25021
Project Title	Prediction and Analysis of liver patient data 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

Suppor Vector Machine	`# Hyperparameter tuning for SVM`
	`svm_params = {`
	`'C': [0.1, 1, 10],`
	`'kernel': ['linear', 'rbf', 'poly'],`
	`'gamma': ['scale', 'auto']`
	`}`
		`print(best_svm)`
		
		`SVC(C=1, gamma='auto')`

Decision Tree	`# Hyperparameter tuning for Decision Tree`
	`dt_params = {`
	`'criterion': ['gini', 'entropy'],`
	`'max_depth': [None, 10, 20, 30],`
	`'min_samples_split': [2, 5, 10]`
	`}`
		`print(best_dt)`
		
		`DecisionTreeClassifier(criterion='entropy', max_depth=30)`

KNN	`# Hyperparameter tuning for KNN`
	`knn_params = {`
	`'n_neighbors': [3, 5, 7, 9],`
	`'weights': ['uniform', 'distance']`
	`}`
		`print(best_knn)`
		
		`KNeighborsClassifier(weights='distance')`

Logistic Regression	`# Hyperparameter tuning for Logistic Regression`
	`lr_params = {`
	`'C': [0.001, 0.01, 0.1, 1, 10],`
	`'solver': ['liblinear', 'saga']`
	`}`
		`print(best_lr)`
		
		`LogisticRegression(C=10, solver='saga')`

Random Forest	`# Hyperparameter tuning for Random Forest`
	`rf_params = {`
	`'n_estimators': [10, 50, 100],`
	`'max_depth': [None, 10, 20],`
	`'min_samples_split': [2, 5, 10]`
	`}`
		`print(best_rf)`
		
		`RandomForestClassifier(max_depth=20, min_samples_split=5)`

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Support Vector Machine	SVM Accuracy: 0.7037037037037037 Classification Report: precision recall f1-score support 1 0.85 0.45 0.59 51 2 0.65 0.93 0.77 57 accuracy 0.70 108 macro avg 0.75 0.69 0.68 108 weighted avg 0.75 0.70 0.68 108 Confusion Matrix: [[23 28] [4 53]]
Decision Tree	Decision Tree Accuracy: 0.7685185185185185 Classification Report: precision recall f1-score support 1 0.69 0.83 0.75 46 2 0.85 0.73 0.78 62 accuracy 0.77 108 macro avg 0.77 0.78 0.77 108 weighted avg 0.78 0.77 0.77 108 Confusion Matrix: [[38 8] [17 45]]

Random Forest	Random Forest Accuracy: 0.8703703703703703 Classification Report: precision recall f1-score support 1 0.82 0.89 0.85 46 2 0.91 0.85 0.88 62 accuracy 0.87 108 macro avg 0.87 0.87 0.87 108 weighted avg 0.87 0.87 0.87 108 Confusion Matrix: [[41 5] [9 53]]
KNN	KNN Accuracy: 0.8055555555555556 Classification Report: precision recall f1-score support 1 0.82 0.70 0.75 46 2 0.80 0.89 0.84 62 accuracy 0.81 108 macro avg 0.81 0.79 0.80 108 weighted avg 0.81 0.81 0.80 108 Confusion Matrix: [[32 14] [7 55]]
Logistic Regression	Logistic Regression Accuracy: 0.7962962962962963 Classification Report: precision recall f1-score support 1 0.76 0.76 0.76 46 2 0.82 0.82 0.82 62 accuracy 0.80 108 macro avg 0.79 0.79 0.79 108 weighted avg 0.80 0.80 0.80 108 Confusion Matrix: [[35 11] [11 51]]

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	The Random Forest 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.