



## **Model Optimization and Tuning Phase Report**

Date	09 July 2024
Team ID	739801
Project Title	Sepsis Survival Minimal Clinical Records
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': [Rome, 10, 20, 30, 40, 50],     'min_samples_split': [2, 5, 10],     'min_samples_leaf': [1, 2, 4] }</pre>	dt train accuracy: 0.9266028881802391 dt test accuracy: 0.9261085233803158 dt train precision: 0.8585929123839606 dt test precision: 0.8576769970776689 dt train recall: 0.9266028881802391 dt test recall: 0.9261085233803158 dt train f1score: 0.8913024242322601 dt test f1score: 0.8905801377924935
Random Forest	<pre># Define the Random Forest classifier rf_classifier = RandomForestclassifier() # Define the hyperparameters and their possible values for tuning param_grid_rf = {</pre>	rf train accuracy: 0.9266028881802391 rf test accuracy: 0.9261085233803158 rf train precision: 0.8585929123839606 rf test precision: 0.8576769970776689 rf train recall: 0.9266028881802391 rf test recall: 0.9261085233803158 rf train f1score: 0.8913024242322601 rf test f1score: 0.8905801377924935







### **Performance Metrics Comparison Report (2 Marks):**

<pre>print(classification_report(y_test,y_test_pred))</pre>
precision recall f1-score support
0 0.00 0.00 0.00 2408
1 0.92 1.00 0.96 29085
accuracy 0.92 31493
Decision Tree macro avg 0.46 0.50 0.48 31493
Weighted avg 0.85 0.92 0.89 31493
<pre>print(confusion_matrix(y_test, y_test_pred))</pre>
[[ 0 2408] [ 0 29085]]



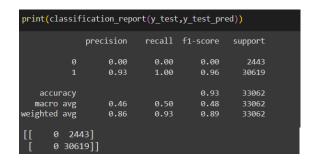


precision recall f1-score support  0 0.00 0.00 0.00 2408 1 0.92 1.00 0.96 29085  accuracy 0.92 31493 macro avg 0.46 0.50 0.48 31493 weighted avg 0.85 0.92 0.89 31493  print(confusion_matrix(y_test, y_test_pred))  [[ 0 2408]
1 0.92 1.00 0.96 29085  accuracy 0.92 31493 macro avg 0.46 0.50 0.48 31493 weighted avg 0.85 0.92 0.89 31493  print(confusion_matrix(y_test, y_test_pred))
Random Forest   macro avg   0.46   0.50   0.48   31493   weighted avg   0.85   0.92   0.89   31493     print(confusion_matrix(y_test, y_test_pred))
[8 242 A
[ 0 29085]]
<pre>print(classification_report(y_test,y_test_pred))</pre>
precision recall f1-score support
0 0.12 0.72 0.20 2408 1 0.96 0.54 0.69 29085
Accuracy 0.56 31493 macro avg 0.54 0.63 0.45 31493 weighted avg 0.89 0.56 0.66 31493
<pre>print(confusion_matrix(y_test, y_test_pred))</pre>
[[ 1724 684] [13243 15842]]
<pre>print(classification_report(y_test,y_pred1))</pre>
precision recall f1-score support 0 0.11 0.73 0.20 2408
1 0.96 0.53 0.69 29085
accuracy 0.55 31493 macro avg 0.54 0.63 0.44 31493 weighted avg 0.89 0.55 0.65 31493
<pre>print(confusion matrix(y_test, y_test_pred))</pre>
[[ 1748 660] [13534 15551]]





# Logistic Regression







# **Final Model Selection Justification (2 Marks):**

Final Model	Reasoning
Random Forest	The decision to select the Random Forest model after applying SMOTE was driven by the need to address class imbalance, leverage the robustness and performance of an ensemble method, gain insights into feature importance, and ensure reliable and correct predictions. This combination aims to enhance the overall predictive capability of the sepsis survival prediction system, contributing to better clinical outcomes.