Model Optimization and Tuning Phase Report

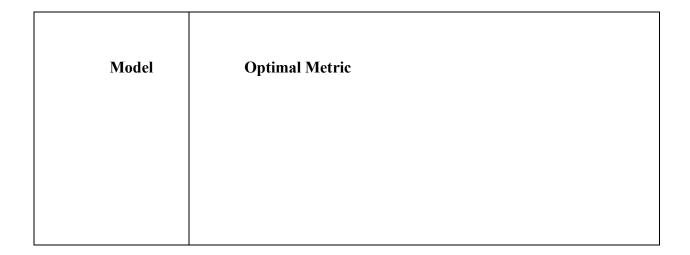| Date	15 July 2024
Team ID	740044
Project Title	One year life expectancy post on Thoracic Surgery using machine learning
Maximum Marks	10 Marks

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

| Random Forest | ```
# Initialize Random forest classifier
rf = RandomForestClassifier(random_state=42)

# Define the parameter grid for hyperparameter tuning with more values
param_grid = {
    'n_estimators': [100, 200, 300, 400, 500],
    'max_features': ['auto', 'sqrt', 'log2'],
    'max_depth': [None, 10, 20, 30, 40, 50],
    'min_samples_split': [2, 5, 10, 15],
    'min_samples_leaf': [1, 2, 4, 6],
    'bootstrap': [True, False]
}
``` | ```
print(f'Accuracy: {accuracy}')
print(f'F1 Score: {f1}')
print('Classification Report:')
print(classification_report(y_test, y_pred))
print('Confusion Matrix:')
print(cm)
``` |

Performance Metrics Comparison Report (2 Marks):

Model	Optimal Metric

Random Forest	
| | ```
Fitting 5 folds for each of 2880 candidates, totalling 14400 fits
Best Parameters: {'bootstrap': True, 'max_depth': None, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 1, 'n_estimators': 200}
Best Score: 0.801355707762521
Accuracy: 0.835164835164835
F1 Score: 0.760158020161076
Classification Report:
              precision    recall  f1-score   support

           0       0.84      1.00      0.91        76
           1       0.00      0.00      0.00        15

    accuracy                           0.84        91
   macro avg       0.42      0.50      0.46        91
weighted avg       0.70      0.84      0.75        91

Confusion Matrix:
[[76  0]
 [15  0]]
``` |

Final Model Selection Justification (2 Marks):

Final Model	Reasoning

Best 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."