1. Why did you choose the particular algorithm?

The Random Forest Classifier was chosen for its robustness against overfitting, ability to provide feature importance scores, versatility with mixed data types, and strong performance with default settings, reducing the need for extensive parameter tuning during initial experimentation.

2.What are the different tuning methods used for the algorithm?

Tuning methods can greatly improve a Random Forest model's performance. Common techniques include Grid Search, which systematically explores hyperparameters, and Random Search, which samples settings from distributions for efficiency. Additionally, employing cross-validation ensures model generalization, while feature selection identifies relevant features to enhance overall performance.

3.Did you consider any other choice of algorithm?Why or why not?

While other algorithms like Logistic Regression, Support Vector Machines (SVM), Gradient Boosting Machines (GBM), XGBoost, and Neural Networks were considered, they each come with challenges such as complexity, high computational cost, and susceptibility to overfitting. Ultimately, Random Forest was chosen for its strong balance of interpretability, performance, and robustness with minimal tuning. It provides reliable results during initial exploration, allowing for further refinements based on its performance without extensive adjustments.

4.What is the accuracy?

Accuracy is the ratio of correctly predicted instances to the total instances in a dataset. It measures a model's overall performance, indicating how well it classifies or predicts outcomes.

5.What are the different types of metrics that can be used to evaluate the model?

Accuracy

Precision

Recall (Sensitivity)

F1 Score

ROC-AUC

Log Loss

Confusion Matrix

Specificity

Matthews Correlation Coefficient (MCC)