1. **Why did you use the particular algorithms?**

**Ans:-** The algorithms used for the problem is Random Forest and XgBoost, as the problem is a multiclass classification problem and they tend to be more efficient for the same.

1. **What are the different tuning methods for the same?**

**Ans:-**

* **Number of Tress(n\_estimators):** Random Forest constructs many decision trees and then combines their forecasts to produce the final result. The number of trees in the forest is an essential hyperparameter that can affect the model's performance. In general, increasing the number of trees improves model accuracy while increasing calculation time. Cross-validation may be used to determine the ideal number of trees for your task.
* **Maximum depth of trees(max\_depth):** The maximum depth of the decision trees governs the model's complexity. A deeper tree can capture more complicated data linkages, but it also raises the danger of overfitting. To discover the best value of max\_depth for your problem, utilise grid search or randomised search.
* **Minimum number of samples required to be at a leaf node (min\_samples\_leaf):** The minimum number of samples necessary to divide a node is controlled by this hyperparameter. A greater min\_samples\_split number can keep the tree from splitting too early and overfitting, but it can also cause underfitting. Cross-validation can be used to determine the best value of min\_samples\_split for your situation.
* **Maximum number of features to consider for each split (max\_features):** At each split, Random Forest chooses a subset of features at random. The max\_features hyperparameter specifies the maximum number of features to take into account at each split. A lower number for max\_features can lessen the correlation across trees and increase the model's generalisation capabilities. To discover the best value of max\_features for your problem, utilise grid search or randomised search.

1. **Do you consider another choice of algorithms? Why or Why Not?**

**Ans:-** I has reached a sufficient accuracy with one of the two models so there was no need of using more algorithms.

1. **What is the accuracy?**

**Ans:-** Maximum accuracy achieved is 94 %.

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

Ans:-

* **Accuracy:** Accuracy is a commonly used metric for classification problems. It measures the proportion of correctly classified samples out of the total number of samples in the dataset. While accuracy is a useful metric, it can be misleading in the case of imbalanced datasets.
* **Precision:** Precision measures the proportion of true positives out of the total number of positive predictions made by the model. It is a useful metric when the goal is to minimize false positives.
* **Recall:** Recall measures the proportion of true positives out of the total number of actual positive samples in the dataset. It is a useful metric when the goal is to minimize false negatives.
* **F1 score:** The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is a useful metric when both false positives and false negatives need to be minimized.
* **Confusion matrix:** A confusion matrix is a table that shows the true positive, true negative, false positive, and false negative predictions made by the model. It is a useful tool for visualizing the performance of the model and calculating various metrics such as accuracy, precision, recall, and F1 score.