## **Sub-questions for Problem 2**

1. Why did you choose the particular algorithm?  
   Ans: I have tried various bagging, and boosting machine learning algorithms. Random Forest classifier gave me 89%, AdaBoost gave 88% accuracy and I achieved 92% accuracy using XGBoost which is the highest accuracy so, XGBoost is the best model, This is a boosting algorithm that reduces the overfitting.
2. What are the different tuning methods used for the algorithm?  
   Ans: I improved my model performance by increasing the number of trees and tried with different values of max\_leaf\_nodes.
3. Did you consider any other choice of the algorithm? Why or why not?  
   Ans: I have tried various other Bagging and boosting machine learning classifiers such as:  
   -Random Forest Random subsamples of data are selected for Random Forest where the growing happens in parallel and overfitting is reduced with the combination of several underfitting features in the algorithm.  
   -Ada Boost One of the many advantages of the AdaBoost Algorithm is it is fast, simple and easy to program. Also, it has the flexibility to be combined with any machine learning algorithm  
   -XGBoost Overfitting is reduced with the help of regularization parameters in XGBoost that helps to select features based on weak and strong features in the decision tree. Optimal values of each leaf are calculated and hence the overall gradient of the tree is given as the output.
4. What is the accuracy?  
   Ans: I have achieved 92% accuracy using XGBoost Model.
5. What are the different types of metrics that can be used to evaluate the model?  
   Ans: The different types of metrics that can be used to evaluate the models are:  
   Accuracy: Accuracy is the number of correctly predicted data points out of all the data points.  
   Precision: Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives).  
   Recall: Mathematically, we define recall as the number of true positives divided by the number of true positives plus the number of false negatives.  
   F1-score: F1 score is defined as the harmonic mean between precision and recall. F1 score is used when dealing with imbalanced data as accuracy is not the preferred metric in these cases.

**Thank You!**