Assignment #3: Ensemble Models for Targeted Marketing Campaign

Business Problem: In this problem, we will use historical data from past customer responses to build a classification model to decide whether a customer is going to accept a marketing offer from the bank or not. We call such person who accepts the offer as "**politically exposed person**" (**PEP**). Rather than doing a mass marketing campaign to all new prospects, we would like to target those that are likely to respond positively to our offer (according to our classification model).

Machine Learning Tasks:

1. Summarize the dataset.
2. Random Forest model.
    a. What is the accuracy of this random forest model on the testing data?
    b. What is the AUC of this random forest model on the testing data?
    c. Is accuracy a good performance metrics or not? If not, why?
    d. What is the training time for the number of trees you built?
    e. How can you speed up the training process by parallelization? Does the model trained via parallelization have the same performance as the one trained using a single core?
3. Gradient boosted decision trees (GBM).
    a. What is the accuracy of this random forest model on the testing data?
    b. What is the AUC of this random forest model on the testing data?
    c. Is accuracy a good performance metrics or not? If not, why?
    d. What is the training time for the number of trees you built?
    e. Does GBM perform better than Random Forest?
    f. If you increase the number of trees in GBM to 500, is the performance on the training data getting better? Is the performance on the testing data getting better as well?

Dataset for Assignment # 3:

Bank Data.csv