Assignment 3: Basic modelling (passing 65%)

Pick any dataset from the list on blackboard

1. Develop a problem statement (real world and machine learning) 10
    a. This is one of the most important skills that a Machine Learning Engineer/Scientist should have. Select a dataset and frame a machine learning problem and then connect this machine learning problem to the real world scenario.
2. Implement the preprocessing and justify the preprocessing steps 20 (without justification 10)
3. Extract features and justify the methods used 20 (without justification 10)
4. Select features and justify the methods used 20 (without justification 10)
5. Implement five out of the following algorithms and justify the choice 30 (without justification 10)
    a. Regression (linear or polynomial)
    b. Additive model
    c. Decision trees
    d. Naïve Bayes
    e. Random forest
    f. SVM with kernels
    g. Neural Network
6. Compare the performance of the five algorithms with respect to your problem, explain the results. 20 (without explanation 10)

Now on the same dataset

7. Implement boosting and bagging with your choice of base models and explain all the steps 20
8. Implement one instance of transfer learning (find a related bigger dataset online) and explain all the steps 20
    a. Explain the bigger dataset with visualization and summary statistics.
9. Compare the performance of the algorithms (basic VS boosting VS bagging VS transfer) with respect to your machine learning problem and explain the results. 20 (without explanation 10)

NOTE 1: There is no word limit here, but the length of the submission will be judged against the material provided. So be careful with what you are writing and how you are writing.

NOTE 2: You can continue with the dataset that you chose in the first assignment or consider choosing a new dataset from BlackBoard