BITS F464 – MACHINE LEARNING

ASSIGNMENT 2 – DECISION TREE AND SUPPORT VECTOR MACHINE

Due Date: Monday, November 27th, 2023, at 23:59 IST

Marks: 15 marks (7.5%)

Assessment type: Written evaluation with model results in a ipynb Notebook.

Overview:

In this assessment, you will have to showcase your knowledge of machine learning techniques learnt so far in the course and their applications. This assessment will allow you to apply your learning and use what you learnt thus far to create models from scratch and compare their performance.

Your answers must be submitted in the same format as that of the supplied ipynb notebook template.

Assessment criteria:

This assessment will measure your ability to:

- Compare, interpret and communicate outputs from various machine learning techniques.

Assessment details:

This assessment aims to help gain hands-on experience by developing and applying appropriate machine learning models. Each scenario carries equal weightage (5 marks each) and assesses the correct application of the respective machine learning technique and the interpretation of the results obtained. Highlight the difference between the below machine Learning techniques by proper understanding of the dataset and code. All implementations need to be done from scratch. Use of ML library functions is not permitted.

1. Decision tree model with entropy implementation
2. Adaboost
3. Multiclass SVM

Dataset:

Title: Communities and Crime

Abstract: The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. The crime dataset contains 128 socio-economic features from the US 1990 Census. The target is the crime rate per community.

Number of Instances: 1994, Number of Attributes: 128

Using the Dataset:

1. Identify the variables that are the most highly correlated with the target
2. Apply either dimensionality reduction or feature selection on the dataset

3. Evaluate multiple algorithms i.e, the algorithms which are mentioned above to predict the crime rate.
4. Compare the performance of each model & identify the best performing one.
5. Present how your model generalizes and performs on unseen data.

Reference for the Dataset: https://archive.ics.uci.edu/dataset/183/communities+and+crime

Notebook requirements:

You are given a sample notebook to complete this assignment. Please enter your code in the appropriate sections of the notebook. For each scenario, there are two main sections: implementation of each machine learning method and your written insights after comparing the two methods.

Evaluation requirements:

For each scenario, your notebook should clearly display the relevant data pre-processing, data modelling with the respective technique and an interpretation of the results obtained.

Submission format:

Use the supplied ipynb notebook for adding your code and insights. Please edit the header markdown text to add your team information (full names and id numbers of each team member).

Your files must be submitted in the following format:

1. Add your team no to the <u>name of the files</u> (For example, **TeamXX_Assignment2.pdf** and **TeamXX_Assignment2.ipynb**). Only one submission per team needs to be submitted, *DO NOT MAKE MULTIPLE SUBMISSIONS.*
2. Please submit **ipynb and PDF version of your code** after running and testing the entire code thoroughly. We will run the notebooks and we should be able to run them seamlessly.
3. Files to be submitted in <u>single folder</u> (named as **TeamXX_Assignment2**) on google classroom:
   a. ipynb notebook
   b. pdf version of ipynb notebook

IMPORTANT NOTE:
   ● *NO MAKEUPS AND LATE SUBMISSIONS WILL BE ACCEPTED AND MARKED.*
   ● *ANY KIND OF PLAGIARISM WILL LEAD TO SEVERE PENALIZATION.*

Contact for clarifications:

In case of any queries, please contact the course's Teaching Assistants (TAs) by email, and any other communication is invalid. It would be best if you wrote a mail to <u>all</u> the following TAs for clarification.

1. S Shashank, p20210412@hyderabad.bits-pilani.ac.in
2. Priya Baju, p20230804@hyderabad.bits-pilani.ac.in
3. Pranjali Attarde, p20220018@hyderabad.bits-pilani.ac.in