Applied Machine Learning

Name: Pragya Shukla

Roll No: MDS202027 (DS 2nd year)

Assignment 1

Description:

In this assignment I have done an experimentation and made a github repository, made five python files, namely evaluate.py,

explore_date.py, prepare_data.py, score.py and train.py.

In prepare_data.py, the aim was to make the data ready for training and required steps like normalisation, etc are performed on it.

In explore_data.py, the task I have done is to plot the different graphs and plots associated with the data to see how the data is. The main aim here was to do some analysis on the data.

In train.py, I have defined the model and trained it on our dataset.

In score.py, the predictions are plotted as per the trained model and stored in a directory.

In evaluate.py, the root mean squared error is computed and displayed.

All the results and files are stored in different directories as required.

Assignment 2

Description:

In this assignment, I have modularized the code and added some description to each method so that we can understand what is the aim of each method.

I have also added a config file that holds all the parameters and hyperparameters that are used in the project, if any changes are to be expected in terms of parameters or hyperparameters used, we can just change them in the config file, they will automatically be reflected in the project.

I have also added git hooks in this part.

Additionally (for bonus), I have added logging support (part of monitoring) to our code, that means we will be able to keep track of where we are and what errors we have encountered in which part of the code.

Basically, there are 5 types of logging support we can add, namely, logging.info, logging.debug, logging.error, logging.warning and logging.critical.

I have added most of them to debug our code while running.

I have also added try exception blocks to stay in touch with errors.

Assignment 3

Description:

This is an attempt to do a bonus assignment, i.e., Transfer Learning using Bert for NLP text classification. In this assignment, I have experimented to identify toxicity in comments. The dataset for this project contains text that may be considered profane, vulgar or profane. I have done the evaluation using ROC curve. I have tried to predict the probability that a comment is toxic or not using transfer learning and BERT. This project includes data pre-processing, splitting

the data to train and test, tokenizing the data, model building and saving the model.