AI-ML Assignment

Dt – 19/10/2022

Below are the steps involved in the assignment:

Please create a virtual environment and install all the dependences libraries mentioned in the requirements.txt

1. Imported the required libraries
2. Loaded the dataset
3. Did the required statistical analysis to understand the dataset better
4. Define a function to do all the required pre processing before vectorizing the required the column
5. In the defined function the following we have done:  
   * 1. removed unwanted symbols,
     2. tokenization,
     3. removed stop words,
     4. removed numbers,
     5. stemming,
     6. positional tagging,
     7. lemmatization.
6. Applied the pre-processing function on the synopsis column.
7. We have used synopsis column as independent variable
8. We have used genre column as dependent variable
9. Then we have built a pipeline to classify the genre
10. We have use the TFIDF vectorizer
11. We have used the Multinomial Naïve Bayes Algorithm for classification
12. Them we have find out the accuracy score in classification
13. For rating prediction, we have used synopsis as the independent variable.
14. We have used the rating as the dependent variable
15. Then we have built a pipeline to predict the rating
16. We have use the TFIDF vectorizer again
17. We have used the KNN regressor Algorithm for regression.
18. Them we have find our the r2 value in regression.