Advanced Machine Learning CS 2490 1 - Spring 2022

Assignment - 1

Instructions

A. Download the datasets from the classroom.
B. Submit code as .ipynb or .py file. If you are submitting multiple files, zip it and rename the file as Firstname_Lastname_AML_HW1.zip.
C. A report in PDF format should be submitted along with the codes. The report will contain details about the approach, results, major findings and discussion on the results.

1) Part 1
    a) Select your dataset.
    b) Convert the data into a two-D binary matrix of user vs product.
    c) Design a recommender system using the 2D matrix. You can use any clustering method. *Autoencoders can be used for clustering*.
    d) Suggest methods for evaluating the correctness of your recommender system.
    e) Present evaluations.
2) Part 2
    a) Each product in the original dataset comes with a product description. Cluster the "product description texts" in order to group products based on their properties – assign cluster ids to the product based on the cluster id. You have to use *neural language models* for this purpose. Indicate in your report which columns you are using for this task, which methods and performance evaluation of the cluster qualities.
    b) Derive a new user–cluster matrix.
    c) Redesign the recommender system
    d) Compare the performance of both recommender systems? (One you made in 1c.)
    e) What kind of insights could you derive about the products and customers from this exercise?
3) Part 3
    a) The datasets also come with a customer-given rating and feedback. Design a BERT based *classification system* for predicting the rating from the feedback text.

4). Part 4

a). Based on the above exercises—share your insights about what drives the rating of the products—what causes more positive or more negative reviews.

b). Based on the insights—come up with a set of recommendations that you can share with the beauty products manufacturer or restaurant owners as areas of improvement.

Timelines—Code and Report Submission for Parts 1, 2 and 3—date: 4th March 2022. In-person presentation will be evaluated on—16th March—a short presentation of 7 to 8 minutes on major findings i.e., part 4 will be made by each student. The findings will have to be based on results of parts 1-3, substantiated by evidence. Any additional

insights about the shortcomings of your analysis are also welcome. The presentation will also have to be submitted.