Assessment-Final Term

Data Warehousing and Data Mining

Marks: 30.00

Submission Deadline: 11/12/2021

Question:

1. Apriori Algorithm usually contains or deals with a large number of transactions. For example, customers buy a lot of goods from a grocery store. By applying this algorithm method, grocery stores can enhance their sales performance and work effectively. Can you consider online grocery store-related datasets as an example and explain the Apriori Algorithm based on your datasets. [Theortical] **Marks: 5**
2. In the data mining area, linear Regression (equations) is used in different real-life applications of including distance and rate problems, pricing problems, calculating dimensions, and mixing different percentages of solutions. Take an experimental dataset and implement the linear Regression algorithm using python language with packages for the prediction task. **Marks: 5**
3. You have learned different types of Clustering algorithms such as K-means clustering, Hierarchical clustering, and Density-based clustering. Also, you are familiar with python based application learning, such as Associate Rule, machine learning algorithms. **Marks: 15**
    - Select three Clustering techniques of your choice (you *need to explain why you chose the particular technique*) and collect a dataset of your choice.

- Apply those techniques to the selected dataset in python/WEKA. Then using python/WEKA, you must be done experiments using the datasets. This process must be well-explained step by step.
- Explain your selected Dataset (E.g., type of attributes, class labels, etc.)
- Your report must be professional and well-explained. If you include graphs or any diagram, then you must explain it scientifically.
- You must provide a valid scientific reason for choosing the technique and the dataset.

4. Explain the Following terms in short: **Marks:5**

   (i) Cross-Validation (ii) Feature selection

Experimental results and Graphs with an analytical report are highly appreciated.

Rubric

Criteria	Excellent (5)	Good (4)	Fair (3)	Poor (1)
Contribution 5%	The contribution of the work/project to the development of scientific concepts is identified and well documented.	Some contribution of the work/project to the development of scientific concepts is identified and documented.	Some contribution of the work/project to the development of scientific concepts is identified but documentation lacks finesse.	No apparent contribution of the work/project to the development of scientific concepts and it has not clearly identified and/or documented.
Methods 10%	Setup was documented completely. Method was also documented completely and accurately, making analysis easy to reproduce.	Setup included descriptive text and diagrams were provided if appropriate. Analysis can be reproduced using the steps provided.	Setup included descriptive text, but diagrams were scarcely used. Hence analysis seemed vague and ambiguous to be replicated.	Description was general or did not include diagrams. Procedure was missing multiple steps. Information provided is not sufficient to replicate experiment.
State-of-Art 5%	Makes the best use of technology and produced a significant result that is likely to have a major impact.	Utilizes the technology but results can be expected to have a modest impact.	Attempts to utilize the technology but results can be expected to have a minor impact.	Does not utilize the technology and the results are obvious or easily anticipated.
Creativity 5%	Deep insight demonstrated and preset a creative solution to the real-life problem.	Some creative solutions have been presented which incrementally improves on previous approaches.	Some creative solutions have been presented but does not improve on previous approaches.	Restated problem and hypothesis. Justified design and methods of experiment.
Conclusion 5%	Restated problem and hypothesis. Justified design and methods of experiment. Findings were discussed in detail. Conclusions directly address hypothesis.	Problem was restated. Statements and conclusions were based on the data collected. Showed a strong relationship between conclusions and hypothesis.	Problem was restated. Statements and conclusions were based on the data collected. But could not develop a strong relationship between conclusions and hypothesis.	Problem was restated. Conclusions were simplistic. No clear relationship between conclusions and hypothesis/objectives.