Assignment

Data Warehousing and Data Mining

Marks: 100.00

Submission Deadline: 13/01/2021

Scenario:

You have learned several classification techniques such as KNN, Naïve Bayes, Decision Tree (ID3, CART).

• You will select one technique of your choice and collect dataset of your choice (Must be explained why you choose particular technique and dataset). Then apply your chosen technique to selected dataset in WEKA. Then using WEKA you must show using the known information, you are able to classify new data points. This process must be well-explained step-by step.

• Dataset must be explained (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 valid scientific reason for choosing the technique and the dataset.

Rubric

Criteria	Excellent (5)	Good (4)	Fair (3)	Poor (1)
Contribution 20%	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 20%	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 20%	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 20%	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 20%	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.