CSCE 474/874: Introduction to Data Mining
Spring 2022

Homework 2 February 8, 2022

Assignment

Implement the *Apriori* algorithm to determine the frequent sets in a dataset and then generate the association rules along with their support and confidence. Inputs to your program must include minimums for support and confidence.

Plot the runtime of your algorithm and the number of rules generated as a function of "minimum support."

Use the same data set to derive association rules in Weka and compare them to those derived from your program.

If you have a dataset from the domain of your project, you are free to use it. You are also encouraged to use the ANES dataset from the last assignment. Use any dataset from Weka datasets otherwise.

Due Date

The assignment is due on February 22 and is worth 100 points.

Handin

Hand in a report along with the listing of your program, the output generated from the run of the test file on Canvas. Make sure that you have uploaded a signed copy of the Contributions form.

Prepare and submit two files as follows:

- Your report named as "Lastname1_Lastname2.pdf" in pdf format. **The signed contributions form should be used as the cover page of your report**.
- A zip file named "Lastname1_Lastname2.zip" that includes everything else (your program, the output generated from the run of the test file, etc.). You must include a README file that describes the usage of your program. Make sure your implementation can successfully execute on the CSE server.