Week 2 Assignment – 40 Points

Questions 1, 2, and 3 should be solved manually. Show all steps of your work.

1. Assume we have 4 items in our set of items you can call them 1, 2 3, and 4
2. How many possibilities are shown on the itemset lattice. Create the itemset lattice.
3. Show the impact of the following assumption on the itemset lattice.

A diagram of a number diagram

Description automatically generated with medium confidenceAssume Itemset {2, 3} is infrequent

A diagram of a network

Description automatically generated

1. Consider the following 10 transaction set.

|  |  |
| --- | --- |
| TID | List of purchased items |
| T1 | A, B, E |
| T2 | B, D |
| T3 | B, C |
| T4 | A, B, D |
| T5 | A, C |
| T6 | B, C |
| T7 | A, C |
| T8 | A, B, C, E |
| T9 | A, B, C |
| T10 | C, D, E |

* 1. What is the set of items? I = {A, B, C, D, E}
  2. Suppose minimum count required is 3, what are set of item set with the equal of greater than 3 count? You should apply Apriori algorithm

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Item set | Count | Item set | Count | Item set | Count | Item set | Count |
| A | 6 | A, B | 4 | A, B, C |  |  |  |
| B | 7 | A, C | 4 |  |  |  |  |
| C | 7 | A, D | 1 |  |  |  |  |
| D | 3 | A, E | 2 |  |  |  |  |
| E | 3 | B, C | 4 |  |  |  |  |
|  |  | B, D | 2 |  |  |  |  |
|  |  | B, E | 2 |  |  |  |  |
|  |  | C, D | 1 |  |  |  |  |
|  |  | C, E | 2 |  |  |  |  |
|  |  | D, E | 1 |  |  |  |  |

1. Use the transaction list and compute the Support, Confidence, and the lift for the following rule: {A, B} => {E}. evaluate the result.

|  |  |
| --- | --- |
| TID | List of purchased items |
| T1 | A, B, E |
| T2 | B, D |
| T3 | B, C |
| T4 | A, B, D, E |
| T5 | A, C |
| T6 | B, C |
| T7 | A, C |
| T8 | A, B, C, E |
| T9 | A, B, C |
| T10 | C, D, E |

Or

Use R for the following question. Submit codes in a separate .txt file

1. Use the market.csv and create a list of rules then:
2. Read the dataset . What is the type of this dataset?
3. Prepare the dataset for applying the algorithm.
4. Apply the algorithm with no parameter specified. Interpret the results
5. Apply the algorithm with the following parameters

(List support = 0.006, confidence = 0.25, minlen = 2)

and prepare a summary of result.

1. Visualization of the rules. Interpret the visualization.
2. Find subsets of rules containing any berry items
3. Visualization of the berry rules. Interpret the visualization
4. Inspect the lift values.
5. writing the rules to a CSV file. Call this file “basketrules.csv”
6. Use basketrules.csv file to create a visualization in Tableau. Interpret the visualization.

Notes:

The groceries.csv is in pure text format and data separated by column you should use the following R code to read this dataset:

basket<-read.transactions(“market.csv”, sep = “,”)

The resulting data in R is in form of a matrix and not a data frame

Required Packages:

* Apriori
* arulesViz
* plotly