**Final Year B. Tech. (CSE) – I: 2022-23**

**5CS462: PE5 - Data Mining Lab**

**Assignment No. 7**

**PRN: 2019BTECS00077 Date:02 oct 2022**

**Full name: Biradar Avinash Vishnu**

**----------------------------------------------------------------------------------------------------------**

**Title:** Design the data analysis tools (GUI) to perform the following task.

1. Implement the Apriori algorithm for generating Association Rules

2. Experiment with different values of support, confidence, and maximum rule length.

3. Tabulate the results containing frequent item sets, total number of rules generated for different support and confidence.

4. Find the interesting rules from above obtained rules using following metrics/measures.

a. Lift b. Chi-Square Test ᵡ 2 c. All\_confidence measure d. Max\_confidence measure e. Kulczynski measure f. Cosine measure

**Objective/Aim:**

1. To implement data analysis tool using python programming language.
2. To design and implement apriori algorithm for generating association rules.

**Introduction:**

Association Rules:

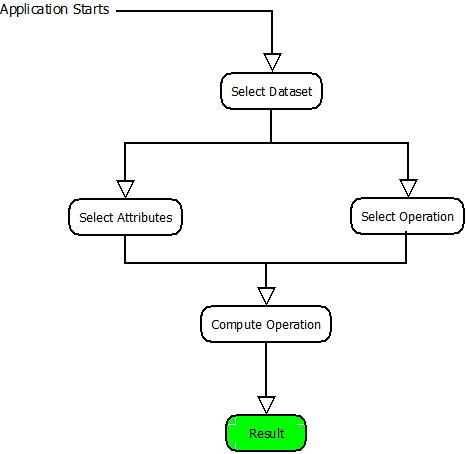
Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occurs in a transaction. A typical example is a Market Based Analysis.

The Association rule is very useful in analyzing datasets. The data is collected using bar-code scanners in supermarkets. Such databases consists of a large number of transaction records which list all items bought by a customer on a single purchase. So the manager could know if certain groups of items are consistently purchased together and use this data for adjusting store layouts, cross-selling, promotions based on statistics.

**Theory:**

* **Support(s) –**  
  The number of transactions that include items in the {X} and {Y} parts of the rule as a percentage of the total number of transaction.It is a measure of how frequently the collection of items occur together as a percentage of all transactions.
* **Support = (X+Y)  total –**  
  It is interpreted as fraction of transactions that contain both X and Y.
* **Confidence(c) –**  
  It is the ratio of the no of transactions that includes all items in {B} as well as the no of transactions that includes all items in {A} to the no of transactions that includes all items in {A}.
* **Conf(X=>Y) = Supp(XY)  Supp(X) –**  
  It measures how often each item in Y appears in transactions that contains items in X also.
* **Lift(l) –**  
  The lift of the rule X=>Y is the confidence of the rule divided by the expected confidence, assuming that the itemsets X and Y are independent of each other.The expected confidence is the confidence divided by the frequency of {Y}.
* **Lift(X=>Y) = Conf(X=>Y)  Supp(Y) –**  
  Lift value near 1 indicates X and Y almost often appear together as expected, greater than 1 means they appear together more than expected and less than 1 means they appear less than expected.Greater lift values indicate stronger association.

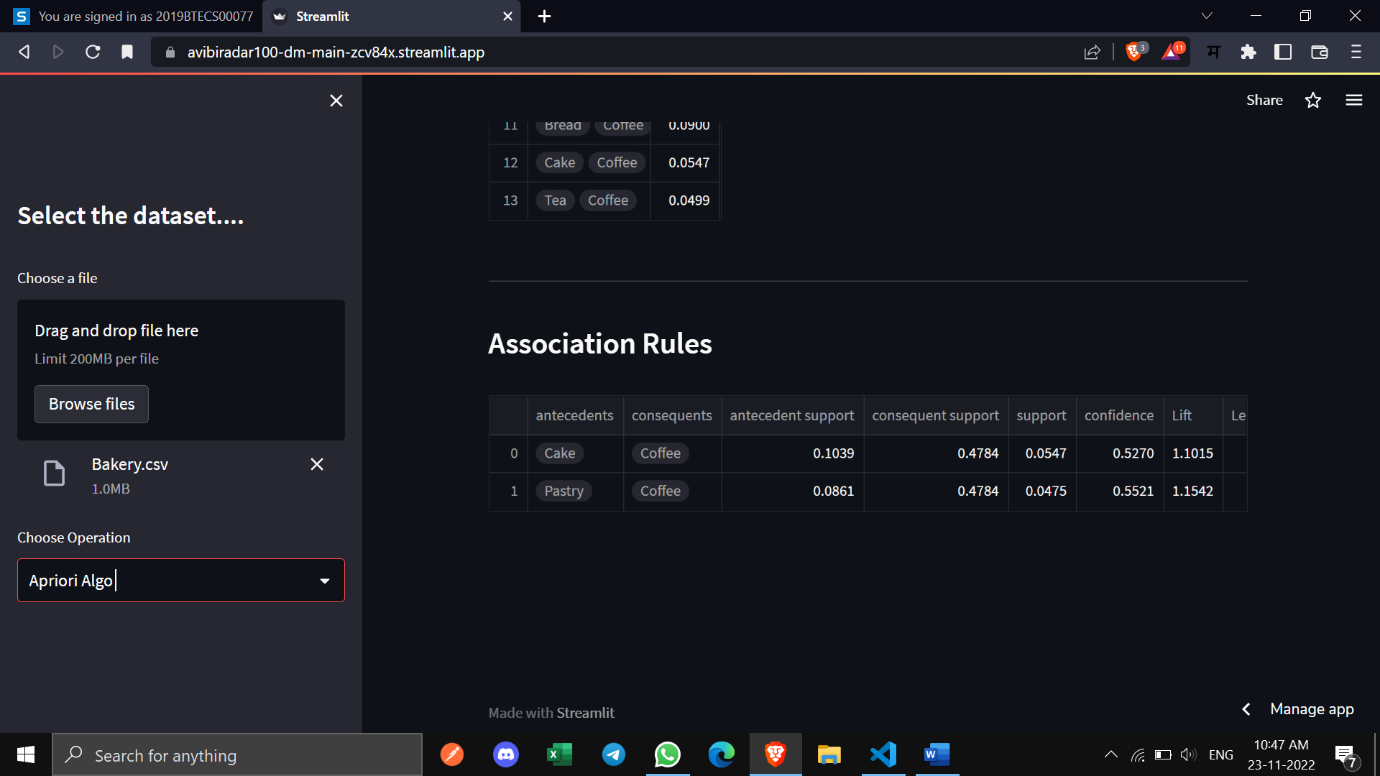
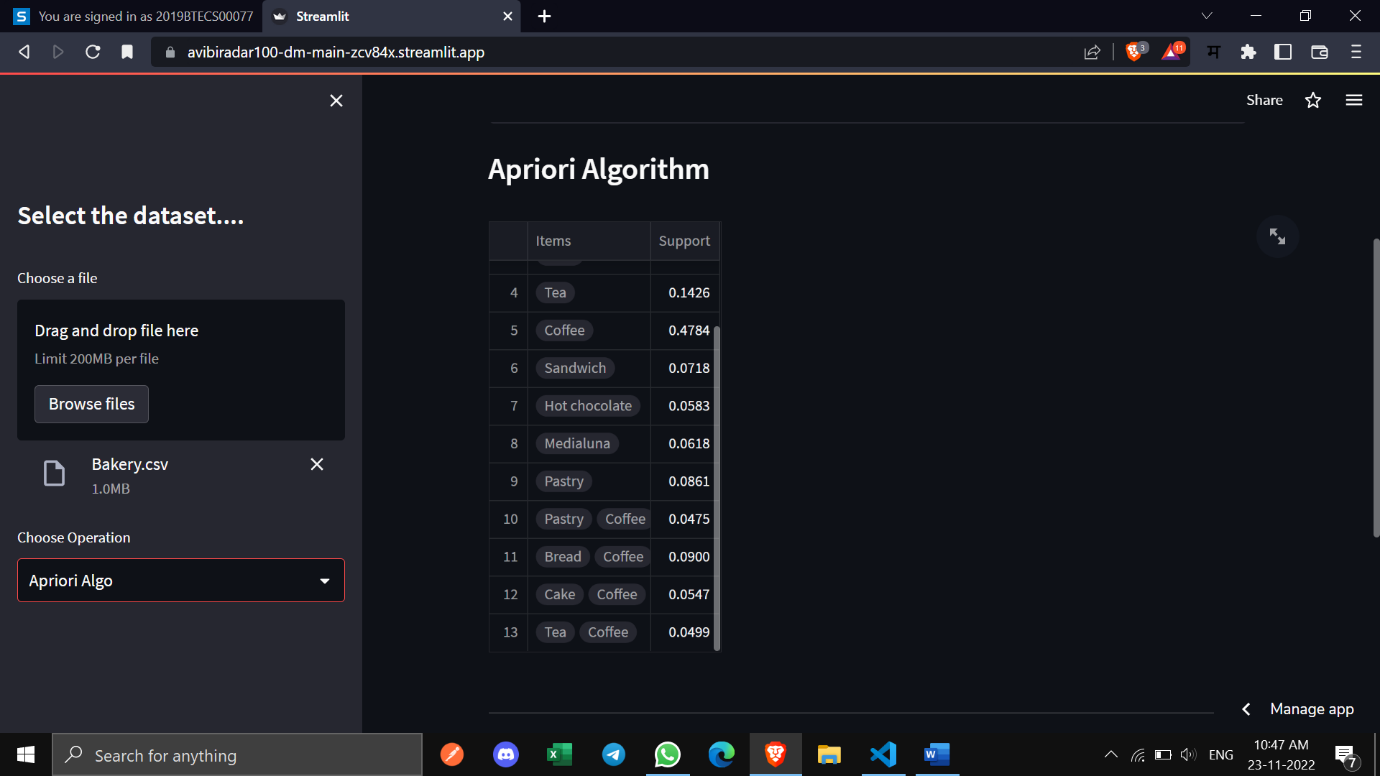
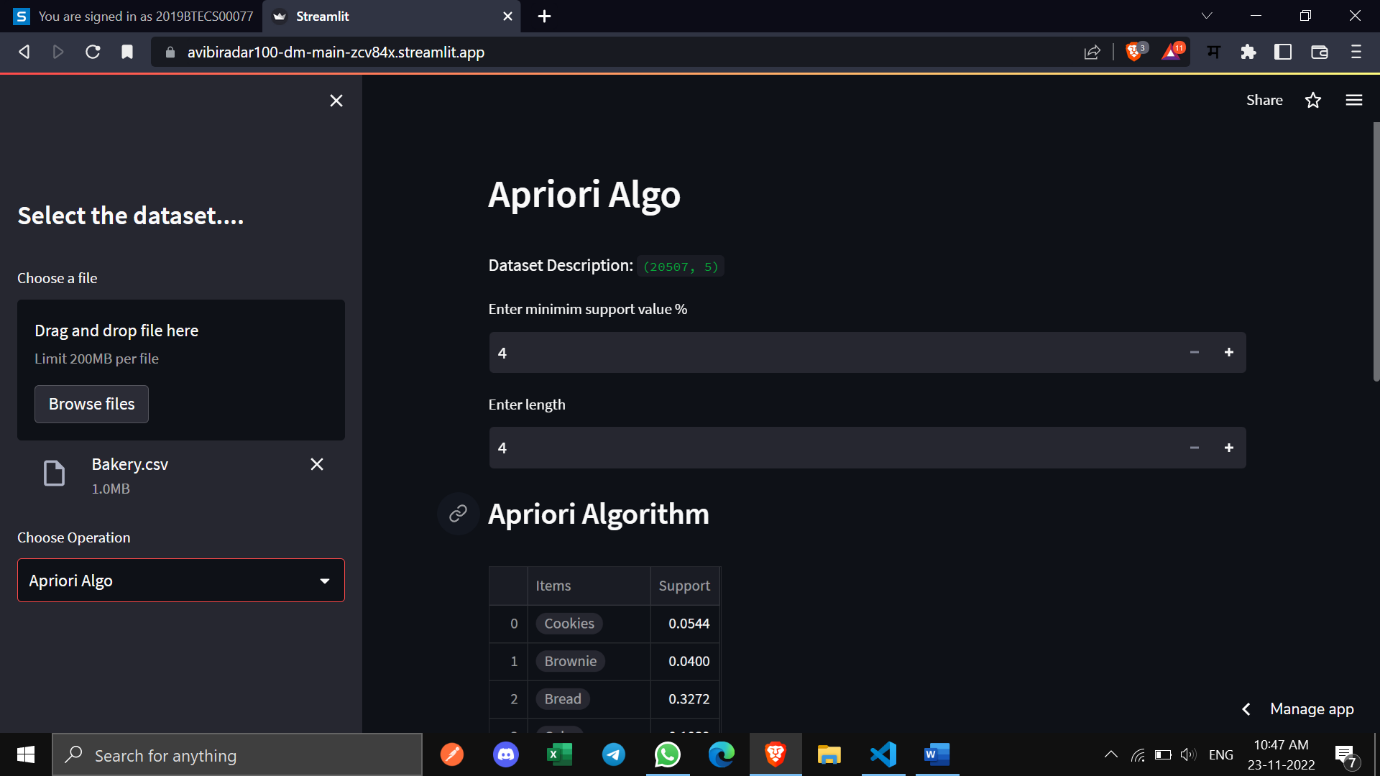
**Functional Block Diagram:**



**Procedure:**

Given problem statement is solved using python programming language and specifically used tkinter module to implement GUI application and pandas module to load .csv file as dataset.

**Actual Experiments/Screenshots:**

****

**Conclusion:**

Successfully implemented data analysis tool (GUI) for implementation of apriori algorithm for association rules.