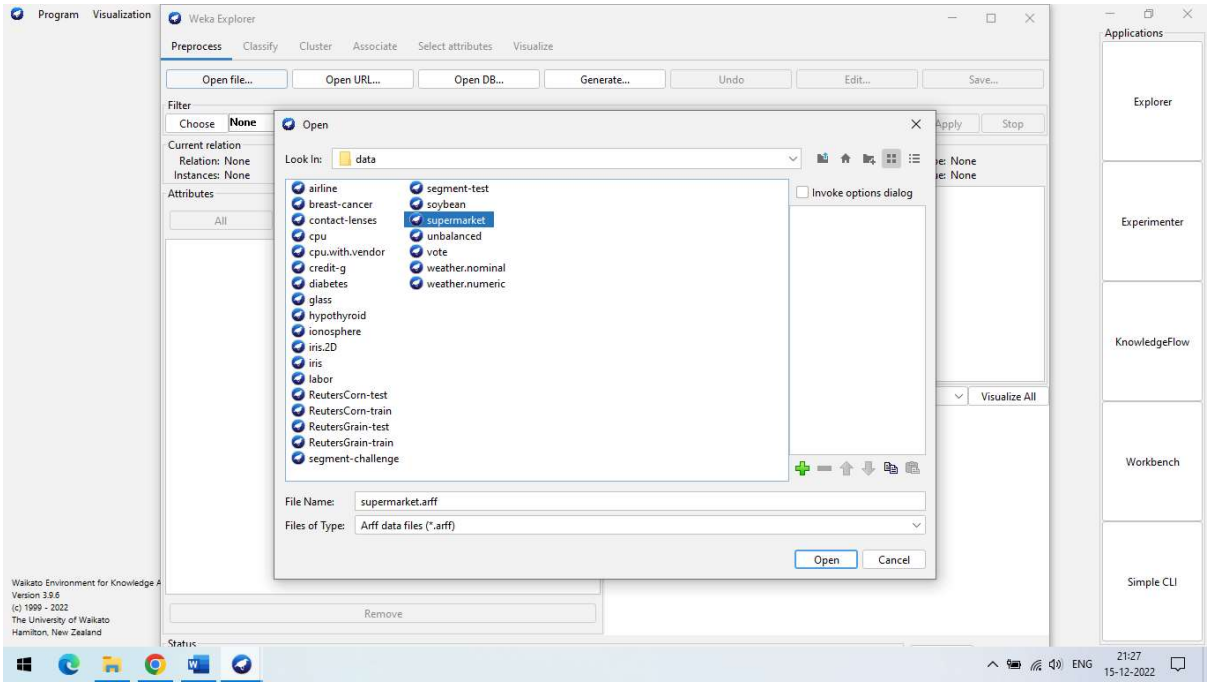
PRACTICAL-8

AIM: Perform Association rule mining using WEKA tool.

Step 1: weka→explorer→open file→supermarket.arff
First of all we are using supermarket.arff dataset



Step 2: Go to **Associate->Choose->Apriori->right click**
Attribute configuration in Weka Explorer as per requirement

NAME

weka.associations.Apriori

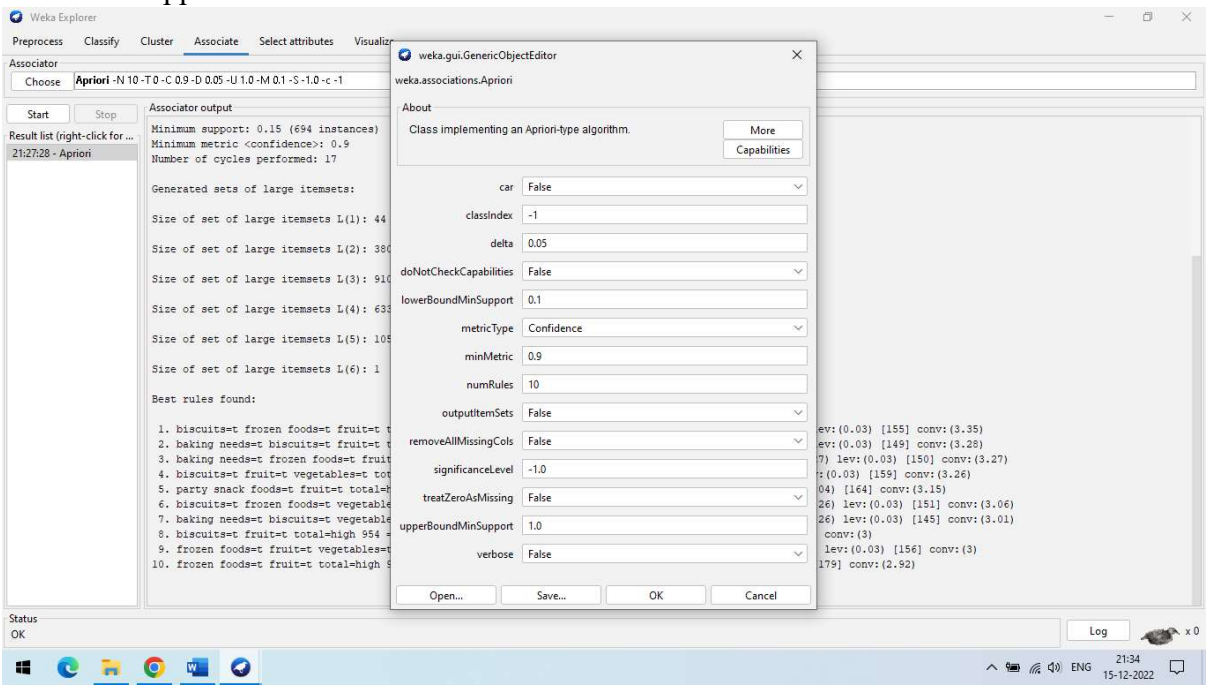
SYNOPSIS

Class implementing an Apriori-type algorithm. Iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence.
The algorithm has an option to mine class association rules. It is adapted as explained in the second reference.
For more information see:
R. Agrawal, R. Srikant: Fast Algorithms for Mining Association Rules in Large Databases. In: 20th International Conference on Very Large Data Bases, 478-499, 1994.
Bing Liu, Wynne Hsu, Yiming Ma: Integrating Classification and Association Rule Mining. In: Fourth International Conference on Knowledge Discovery and Data Mining, 80-86, 1998.

OPTIONS

minMetric -- Minimum metric score. Consider only rules with scores higher than this value.
verbose -- If enabled the algorithm will be run in verbose mode.
numRules -- Number of rules to find.
lowerBoundMinSupport -- Lower bound for minimum support.
classIndex -- Index of the class attribute. If set to -1, the last attribute is taken as class attribute.
outputItemSets -- If enabled the itemsets are output as well.
car -- If enabled class association rules are mined instead of (general) association rules.

doNotCheckCapabilities -- If set, associator capabilities are not checked before associator is built (Use with caution to reduce runtime).
removeAllMissingCols -- Remove columns with all missing values.
Significancelevel -- Significance level. Significance test (confidence metric only).
Treatzeroasmissing -- If enabled, zero (that is, the first value of a nominal) is treated in the same way as a missing value.
Delta -- Iteratively decrease support by this factor. Reduces support until min support is reached or required number of rules has been generated.
Metrictype -- Set the type of metric by which to rank rules. Confidence is the proportion of the examples covered by the premise that are also covered by the consequence (Class association rules can only be mined using confidence). Lift is confidence divided by the proportion of all examples that are covered by the consequence. This is a measure of the importance of the association that is independent of support. Leverage is the proportion of additional examples covered by both the premise and consequence above those expected if the premise and consequence were independent of each other. The total number of examples that this represents is presented in brackets following the leverage. Conviction is another measure of departure from independence. Conviction is given by P(premise)P(!consequence) / P(premise, !consequence).
upperBoundMinSupport -- Upper bound for minimum support. Start iteratively decreasing minimum support from this value.



Step 3: Go to **Associate->Choose->Apriori->start**
The Apriori Algorithm is an influential algorithm for mining frequent itemsets for boolean association rules. Some key concepts for Apriori algorithm are :
• Frequent Itemsets: The sets of item which has minimum support (denoted by Li for ith-Itemset).
• Apriori Property: Any subset of frequent itemset must be frequent.
• Join Operation: To find Lk , a set of candidate kitemsets is generated by joining Lk-1 with itself.

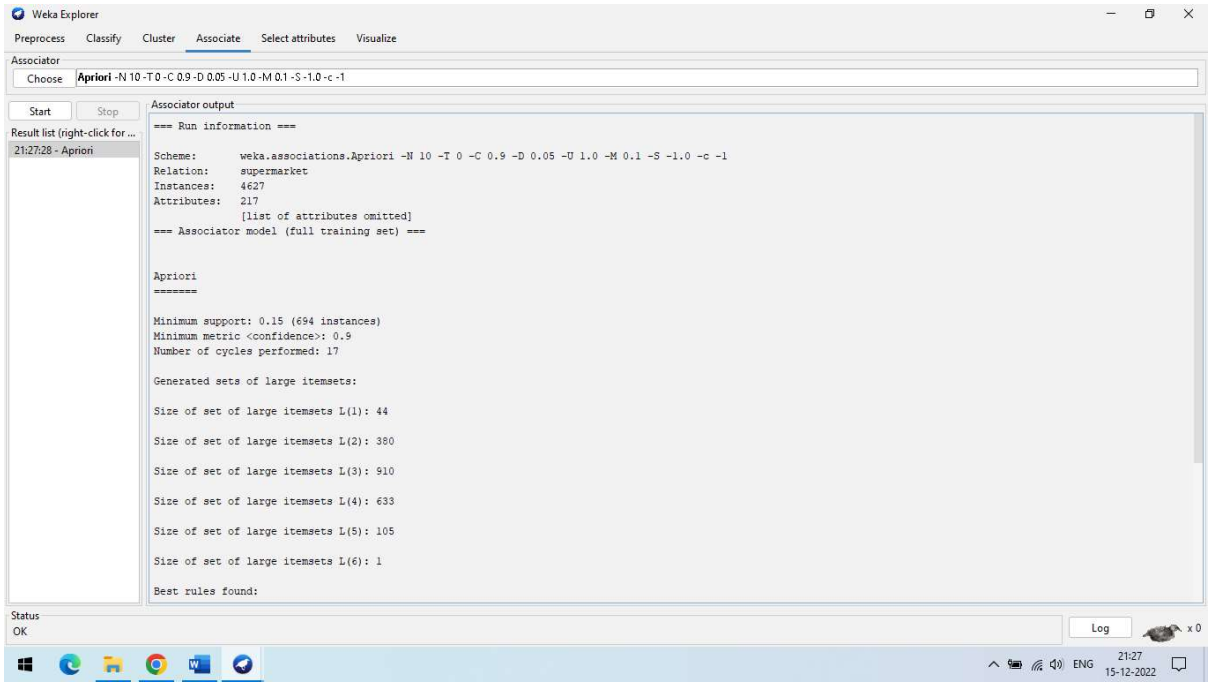
CAPABILITIES

Class -- Binary class, Missing class values, No class, Nominal class

Attributes -- Binary attributes, Empty nominal attributes, Missing values, Nominal attributes, Unary attributes

Additional

Minimum number of instances: 1



Apriori is the simplest algorithm which is used for mining of frequent patterns from the transaction database. The purpose of reducing the number of scans of database to extract frequent item set will be resolved in future due to our work is in progress for the same.

Goal: finding inherent regularities in data