In this report, the rule generation algorithm i.e., Association rule and Apriori algorithm are applied to the groceries dataset to find the insight of the grocery store items sales using data mining techniques. Firstly, the applied algorithm is explained and after that the result of algorithm are discussed.

Frequent itemset generation is a technique used in data mining and machine learning to identify sets of items that frequently occur together in a dataset. This technique is often used in market basket analysis, where the goal is to identify items that are commonly purchased together in a retail setting.

Rule generation is the process of automatically creating rules from data. These rules can be used for a variety of tasks such as classification, prediction, and decision making. There are several techniques used in rule generation, including decision tree induction, association rule mining, and Bayesian networks. These techniques can be applied to both structured and unstructured data, and can be used to generate a wide range of rules, from simple if-then statements to more complex relationships.

One Hot Encoding:

One-hot encoding is a process of converting a categorical variable into a vector of binary values. Each unique value in the categorical variable is represented as a binary vector, with a 1 in the position corresponding to the value and 0s in all other positions. This is useful in machine learning and statistics when working with categorical data. It is also useful in neural networks, to convert categorical data into a format that can be input into the network.

Association Rule Generation

Association rule generation is a specific technique used to discover relationships between items in a dataset. The technique is typically used in market basket analysis, where the goal is to find items that are frequently bought together. The Apriori algorithm is one of the most widely used algorithm for association rule generation.

Apriori Algorithm

The Apriori algorithm works by first identifying all items that appear together in a transaction with a minimum support threshold. The support threshold is a measure of how frequently an item or a combination of items appears in the dataset. Once all the items that meet the minimum support threshold have been identified, the algorithm then generates rules that describe the relationships between those items.

The generated rules take the form of "if-then" statements, where the "if" part represents a combination of items that frequently appear together, and the "then" part represents another item that also frequently appears with those items. The rules are ranked according to a measure of their strength, such as confidence or lift.

Association rule generation and Apriori algorithm are widely used in various fields such as retail, marketing, and healthcare, to gain insights on customer behavior, product placement, and stock management.

Dataset:

The Dataset chosen is of grocery store transaction record containing the following columns

- Member Number
- Date
- Item Description

It should be noted that if a customer bought multiple products on same day, we will consider it one transaction.

Analysis:

The analysis shows that the top ten item sold in grocery store include whole milk, other vegetables, rolls/buns, soda, yogurt, root vegetables, tropical fruit, bottled water, sausage and citrus fruit. The whole milk was sold with frequency of nearly 2500 with highest among all. The detail of the analysis is shown in figure 1

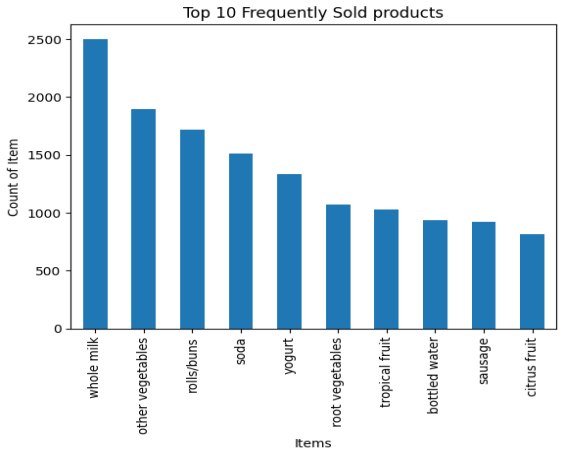

Figure 1 Top 10 items sold

Results:

In this report, two algorithms are applied on above explained dataset i.e., association rule and apriori algorithm to find out the frequent used items in the grocery store. In case of apriori, the result shows that “tropical fruit” item is most frequently used item in grocery store and other are “whole milk” and “pip fruit” with on second and third number respectively. However, in association rule approach results shows that most conviction value is of the bottled beer and whole milk as antecedents and consequents. In second, the sausages and whole milk have conviction value and results clearly shows that whole milk is purchased with different items like bottled beer, newspaper, sausages and domestic eggs as their antecedents. It was also noted that total 74 rules were made to elaborate the relations.s