D212 Market Basket Analysis

The purpose of this Data Mining Report is to identify key relationships among variables in the dataset. One question that can translate to the real-world organization would be to find relationships among the variables that lead to someone leaving the cable company they are currently contracted with. One goal is to find these relationships to see how they can assist the company in turning some profit to present to the stakeholders.

Market Basket Analysis is very different compared to other models. Market Basket Analysis is dependent on the dataset having columns where all the values are either True or False or 1’s and 0’s. This is different than most models that allow features to have continuous or ordinal data. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows retailers to identify relationships between the items that people buy (Li, 2017). The expected outcomes consist of two outputs called Antecedents and Consequents. These are associated with “If this then that” scenarios or rules. One example of a transaction in the dataset would be: If a customer has a yearly contract instead of paying month to month then they don’t leave the company. One assumption of Market Basket Analysis is that the Apriori algorithm works with the assumption that any subset of a frequent itemset must be frequent (Vadakkanmarveettil, 2021).

The Apriori algorithm was able to process the data after it had been converted to Boolean values. Once the data was processed then the code was able to generate the association rules. Attached are screenshots of the code and the output of rules.

**Table

Description automatically generated**

Once the Apriori algorithm was able to run, specific parameters were applied to see what rules appeared to show the strongest relations. Attached is a screenshot for reference. The rules were filtered to have an antecedent greater that .01, a support greater than .009, confidence greater than .85, and lift greater than 1.00. Once these filters were applied the top three rules that populated were if a person lived in a rural area then they were signed up with the phone plan, if a person lived in a suburban area then they were signed up with the phone plan, and if a person was divorced then then they were signed up with the phone plan.

**Table

Description automatically generated**

In summary, the dataset provided very strong rules where the consequent was typically whether a customer was signed up with the phone plan. Since our goal was to see if there was any relationship that showed whether a person left the cable company or not, the associated rules were ran again but this time only focusing on the consequent being whether a person left the cable company. Attached is a screenshot that shows the output of this analysis. Graphical user interface, application

Description automatically generated

What this table is able to show us is that the top three rules are if a person doesn’t stream movies and they are on a two year contract then they don’t leave the cable company, if a person is on a one year plan and they don’t stream movies then they wont leave the cable company, and if the person is on a two year contract and they don’t stream television then they won’t leave the health plan.

The significance of support was that it was positive but the key findings were that confidence and lift show a strong relationship. Lift has a value greater 1. When lift has a value greater than 1 then this implies the relationship is not due to chance. Confidence tells us what the likelihood was that both the antecedent and the consequent would occur. The analysis shows that the confidence was almost at 100%.

The practical significance of the findings would be that in all three rules the main components were that people were on a yearly contract and that they did not stream any kind of service through their internet. This can imply that they are using a different company as their internet provider. A recommended course of action would be to identify the people that fall in the three demographics and promote our internet services to ensure they began to stream their services with us.

# Bibliography

Li, S. (2017, 09 24). *A Gentle Introduction on Market Basket Analysis — Association Rules* . Retrieved from Towards Data Science: https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

Vadakkanmarveettil, J. (2021, 04 05). *Market Basket Analysis – A Complete Overview (2021)* . Retrieved from jigsawacademy.com: https://www.jigsawacademy.com/market-basket-analysis/