D212: Data Mining II: Task 3

Performance assessment

Part A: Research Question

1. A question that could be asked of this dataset is to figure out which medications typically are bought in tandem with each other, and would this help us determine what kind of diseases are being treated or if overprescribing is happening?
2. The main goal of this market basket analysis is to see what medications are bought with each other most often, which would help us determine the kinds of diseases that seem to pair up and how medical communities would be able to address these concerns to help reduce these medical issues.

Part B: Market Basket Justification

1. The main goal of a market basket analysis is to determine items within a dataset that appear to show up together. MBA determines how related certain items are and the frequency to which they occur together to help determine the directness of their relationship. It’s a useful analysis for finding patterns within a dataset.
2. I have attached my code to see all the sets that could be generated, but an example of a set would simply be a set that had a patient with the combined set of acetaminophen and alprazolam.
3. The main assumption that is used for market basket analysis is that each transaction is its own set. Each set can be treated on its own and compared to the other sets and transactions to find patterns.

Part C: Data Preparation and Analysis

1. I cleared out all the null rows and had the dataset down to 7501 records. Attached is the cleaned dataset.
2. See attached for error free code
3. See attached for support, lift, and confidence values.
4. See attached for the rules that were generated from the Apriori algorithm.

Part D: Data Summary and Implications

1. Support tells you how popular a certain itemset is. This is figured out by taking the number of times the item shows up in the dataset. Lift is telling the analyst how likely a secondary criteria/purchase is made when a primary criteria/purchase is made. Lift makes sure to control for the secondary criteria based on how popular it is. Confidence is how likely the 2 items are meant to be grouped/bought together. This checks the likelihood of a specific secondary item accompanying a primary item.
2. The practical findings from a market basket analysis of this kind of dataset can really help figure out what kind of medications are being bought together. This would have enormous benefits for any medical organization as it would help understand what symptoms are commonly associated with each other as well as possibly finding out if ill effects could arise from the interaction of various medications together.
3. For my question I wanted to know what medicines tend to get paired up with other medicines and see if that could help us predict other ailments that a patient might be having. Unfortunately, this dataset doesn’t provide us with very much confidence in the prediction of associated goods in this market basket analysis. The best result we got was that there is a roughly 50% confidence that people who get carvedilol will also get amlodipine. This relationship makes since because carvedilol is a high blood pressure medication and amlodipine is for high blood pressure and chest pain. This makes since that they would be rather related to each other if they tend to treat similar conditions. This analysis could be improved by getting this data associated with patient conditions and perhaps performing a market basket analysis on the medications along with different types of diseases and ailments that patients have.

Part E: Video

Panopto Video: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=8ebb3be7-6b46-4160-a2b5-ae290183ea37>

Sources:

<https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-market-basket-analysis/> A Comprehensive Guide on Market Basket Analysis, Amruta Kadlaskar

<https://pbpython.com/market-basket-analysis.html> Introduction to Market Basket Analysis in Python, Chris Moffitt

<https://towardsdatascience.com/apriori-association-rule-mining-explanation-and-python-implementation-290b42afdfc6> Apriori: Associaton Rule Mining In-depth Explanation and Python Implementation, Chonyy

<https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html> Association Rules and the Apriori Algorithm: A Tutorial, Annalyn Ng

<https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00> Market Basket Analysis 101: Key Concepts, Susan Currie Sivek, Ph.D.

<https://www.softwaretestinghelp.com/apriori-algorithm/> Apriori Algorithm In Data Mining: Implementation With Examples