­

**Performance Assessment for D212: Data Mining II  
Task 3: Market Basket Analysis**

Christopher Fischer - 011933891

College of Information Technology, Western Governors University

October 1, 2024

Performance Assessment for D212: Data Mining II – Task 3

This document contains the tasks and outputs required for the “OFM4 TASK 3: ASSOCIATION RULES AND LIFT ANALYSIS” assessment. All work is original to the author unless otherwise indicated by a citation.

# A - Research Question and Goal

This study aims to answer the following question: Can we use market basket analysis on pharmacy transaction data and identify medications more likely to be ordered together? The primary goal of this study is to increase efficiency in our pharmacy operations by optimizing the arrangement of medicines in inventory.

# B – Method Justification

Market Basket Analysis, hereafter MBA, is a machine learning method by which customer transaction data is mined to find patterns among items appearing together in the transactions. (Pai, 2024) The algorithm works by making pairs of antecedents/consequents; in other words, milk was found (antecedent) in a transaction with eggs (consequent). These pairs are then ranked by how strong and frequent the associations are in each pairing.

This algorithm could be easily “brute forced” if the number of unique items in the dataset is quite small. However, the number of possible associations grows exponentially as that number increases. Therefore, additional methods to reduce the complexity of the model are essential. This study will use the Apriori algorithm to reduce the number of antecedent/consequent pairs that must be considered. Apriori follows the logical rule that if a set of items occurs infrequently, all supersets containing those items will also occur infrequently. (Noble, 2024) This rule is the key assumption in the MBA algorithm. If there is insufficient variability in the transaction data, either because the dataset is too small or non-representative of actual behavior, then the output of the MBA may not be suitable for the intended purpose.

Our team expects to identify a set of medications that are frequently ordered together. This will allow our inventory management team to rearrange product placement in our pharmacies to reduce the time pharmacists and technicians spend filling orders by minimizing unnecessary movement within the pharmacy.

This is an example of one transaction provided in the dataset for this study.

|  |  |
| --- | --- |
| **Column** | **Value** |
| Presc01 | abilify |
| Presc02 | atorvastatin |
| Presc03 | folic acid |
| Presc04 | naproxen |
| Presc05 | losartan |
| Presc06 | NaN |
| Presc07 | NaN |
| Presc08 | NaN |
| Presc09 | NaN |
| Presc10 | NaN |
| Presc11 | NaN |
| Presc12 | NaN |
| Presc13 | NaN |
| Presc14 | NaN |
| Presc15 | NaN |
| Presc16 | NaN |
| Presc17 | NaN |
| Presc18 | NaN |
| Presc19 | NaN |
| Presc20 | NaN |

# C – Data Preparation and Analysis

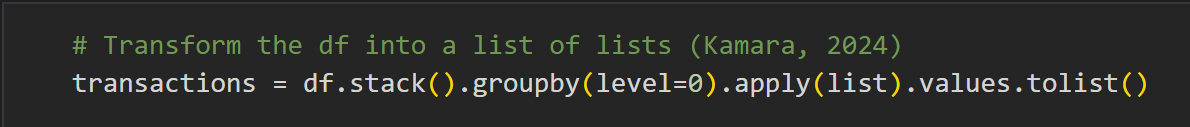
## C1- Data Preparation

Transactional data was provided to our team in the *medical\_market\_basket.csv* input file. This file required a few steps to prepare it for MBA.

A screenshot of a computer

Description automatically generated

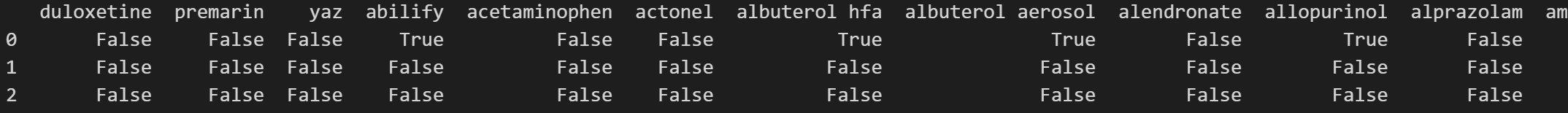
The input file contains alternating rows of nulls between each usable data row. These were removed.



Transactions were transformed into a list of lists for use with the TransactionEncoder.

A screen shot of a computer code

Description automatically generated



Transactions were transformed into a data frame of one-hot encoded transactions.

The prepared data set will be uploaded with this report. See file *medical\_market\_basket\_prepared.csv*.

## C2 – Association Rule Generation

The following steps were taken to apply the Apriori algorithm and generate a set of association rules.

A screenshot of a computer

Description automatically generated

A data frame of frequent item sets was created with a minimum support level of 0.02 and a maximum number of items in each set of two.

A screen shot of a computer code

Description automatically generated

Association rules were generated using the item sets created in the previous step. Lift was used as the metric for evaluation with a minimum value of 1.0. Then, antecedent and consequent item sets were converted to strings for easier visualization. Lastly, the resulting rules were exported to *medical\_market\_basket\_rules.csv*, which is included with this submission.

## C3 – Association Rules Table

The resulting rules are shown here, sorted by confidence in descending order.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **antecedents** | **consequents** | **support** | **lift** | **confidence** |
| metformin | abilify | 0.023 | 1.915 | 0.456 |
| glipizide | abilify | 0.028 | 1.758 | 0.419 |
| lisinopril | abilify | 0.041 | 1.748 | 0.417 |
| lisinopril | carvedilol | 0.039 | 2.291 | 0.399 |
| fenofibrate | abilify | 0.020 | 1.654 | 0.394 |
| clopidogrel | abilify | 0.023 | 1.594 | 0.380 |
| metoprolol | abilify | 0.036 | 1.572 | 0.375 |
| atorvastatin | abilify | 0.048 | 1.554 | 0.370 |
| amphetamine salt combo | abilify | 0.024 | 1.497 | 0.357 |
| doxycycline hyclate | abilify | 0.034 | 1.489 | 0.355 |
| glipizide | carvedilol | 0.023 | 2.000 | 0.348 |
| naproxen | abilify | 0.020 | 1.443 | 0.344 |
| carvedilol | abilify | 0.060 | 1.439 | 0.343 |
| dextroamphetamine xr | abilify | 0.027 | 1.421 | 0.339 |
| citalopram | amphetamine salt combo xr | 0.029 | 1.838 | 0.330 |
| amlodipine | abilify | 0.024 | 1.385 | 0.330 |
| diazepam | abilify | 0.053 | 1.348 | 0.321 |
| levofloxacin | abilify | 0.020 | 1.342 | 0.320 |
| cialis | abilify | 0.024 | 1.316 | 0.314 |
| amphetamine salt combo | carvedilol | 0.021 | 1.758 | 0.306 |
| amlodipine | carvedilol | 0.021 | 1.704 | 0.297 |
| metoprolol | carvedilol | 0.028 | 1.679 | 0.292 |
| amphetamine salt combo xr | abilify | 0.051 | 1.189 | 0.283 |
| citalopram | abilify | 0.024 | 1.174 | 0.280 |
| atorvastatin | carvedilol | 0.035 | 1.572 | 0.274 |
| doxycycline hyclate | carvedilol | 0.025 | 1.522 | 0.265 |
| citalopram | glyburide | 0.022 | 1.476 | 0.252 |
| abilify | carvedilol | 0.060 | 1.439 | 0.251 |
| atorvastatin | diazepam | 0.032 | 1.513 | 0.248 |
| metoprolol | atorvastatin | 0.024 | 1.910 | 0.248 |
| citalopram | carvedilol | 0.021 | 1.414 | 0.246 |
| metoprolol | diazepam | 0.023 | 1.468 | 0.241 |
| diazepam | carvedilol | 0.039 | 1.374 | 0.239 |
| atorvastatin | amphetamine salt combo xr | 0.031 | 1.322 | 0.238 |
| lisinopril | diazepam | 0.023 | 1.433 | 0.235 |
| doxycycline hyclate | amphetamine salt combo xr | 0.022 | 1.272 | 0.229 |
| metoprolol | amphetamine salt combo xr | 0.022 | 1.269 | 0.228 |
| carvedilol | diazepam | 0.039 | 1.374 | 0.225 |
| carvedilol | lisinopril | 0.039 | 2.291 | 0.225 |
| lisinopril | atorvastatin | 0.022 | 1.728 | 0.224 |
| abilify | diazepam | 0.053 | 1.348 | 0.221 |
| losartan | glyburide | 0.029 | 1.263 | 0.216 |
| abilify | amphetamine salt combo xr | 0.051 | 1.189 | 0.214 |
| glyburide | amphetamine salt combo xr | 0.036 | 1.185 | 0.213 |
| doxycycline hyclate | glyburide | 0.020 | 1.239 | 0.212 |
| diazepam | glyburide | 0.034 | 1.228 | 0.210 |
| carvedilol | amphetamine salt combo xr | 0.037 | 1.167 | 0.210 |
| carvedilol | atorvastatin | 0.035 | 1.572 | 0.204 |
| amphetamine salt combo xr | carvedilol | 0.037 | 1.167 | 0.203 |
| diazepam | amphetamine salt combo xr | 0.033 | 1.127 | 0.203 |
| amphetamine salt combo xr | glyburide | 0.036 | 1.185 | 0.203 |
| abilify | atorvastatin | 0.048 | 1.554 | 0.201 |
| glyburide | diazepam | 0.034 | 1.228 | 0.201 |
| losartan | carvedilol | 0.027 | 1.153 | 0.201 |
| diazepam | atorvastatin | 0.032 | 1.513 | 0.196 |
| losartan | amphetamine salt combo xr | 0.025 | 1.072 | 0.193 |
| amphetamine salt combo xr | diazepam | 0.033 | 1.127 | 0.185 |
| atorvastatin | glyburide | 0.024 | 1.071 | 0.183 |
| atorvastatin | metoprolol | 0.024 | 1.910 | 0.182 |
| losartan | diazepam | 0.023 | 1.084 | 0.178 |
| abilify | lisinopril | 0.041 | 1.748 | 0.172 |
| amphetamine salt combo xr | atorvastatin | 0.031 | 1.322 | 0.171 |
| atorvastatin | lisinopril | 0.022 | 1.728 | 0.170 |
| glyburide | losartan | 0.029 | 1.263 | 0.167 |
| amphetamine salt combo xr | citalopram | 0.029 | 1.838 | 0.160 |
| carvedilol | metoprolol | 0.028 | 1.679 | 0.160 |
| carvedilol | losartan | 0.027 | 1.153 | 0.152 |
| abilify | metoprolol | 0.036 | 1.572 | 0.150 |
| carvedilol | doxycycline hyclate | 0.025 | 1.522 | 0.145 |
| diazepam | losartan | 0.023 | 1.084 | 0.143 |
| amphetamine salt combo xr | losartan | 0.025 | 1.072 | 0.142 |
| abilify | doxycycline hyclate | 0.034 | 1.489 | 0.141 |
| diazepam | lisinopril | 0.023 | 1.433 | 0.141 |
| diazepam | metoprolol | 0.023 | 1.468 | 0.140 |
| glyburide | atorvastatin | 0.024 | 1.071 | 0.139 |
| carvedilol | glipizide | 0.023 | 2.000 | 0.132 |
| glyburide | citalopram | 0.022 | 1.476 | 0.129 |
| carvedilol | citalopram | 0.021 | 1.414 | 0.123 |
| carvedilol | amlodipine | 0.021 | 1.704 | 0.122 |
| amphetamine salt combo xr | metoprolol | 0.022 | 1.269 | 0.121 |
| amphetamine salt combo xr | doxycycline hyclate | 0.022 | 1.272 | 0.121 |
| carvedilol | amphetamine salt combo | 0.021 | 1.758 | 0.120 |
| glyburide | doxycycline hyclate | 0.020 | 1.239 | 0.118 |
| abilify | glipizide | 0.028 | 1.758 | 0.116 |
| abilify | dextroamphetamine xr | 0.027 | 1.421 | 0.115 |
| abilify | citalopram | 0.024 | 1.174 | 0.102 |
| abilify | amphetamine salt combo | 0.024 | 1.497 | 0.102 |
| abilify | cialis | 0.024 | 1.316 | 0.101 |
| abilify | amlodipine | 0.024 | 1.385 | 0.099 |
| abilify | metformin | 0.023 | 1.915 | 0.097 |
| abilify | clopidogrel | 0.023 | 1.594 | 0.096 |
| abilify | levofloxacin | 0.020 | 1.342 | 0.085 |
| abilify | fenofibrate | 0.020 | 1.654 | 0.084 |
| abilify | naproxen | 0.020 | 1.443 | 0.084 |

Table - Association Rules

## C4 – Top Three Association Rules

These are the top three association rules generated by this study.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **antecedents** | **consequents** | **support** | **lift** | **confidence** |
| metformin | abilify | 0.023 | 1.915 | 0.456 |
| glipizide | abilify | 0.028 | 1.758 | 0.419 |
| lisinopril | abilify | 0.041 | 1.748 | 0.417 |

Table - Top Three Association Rules

These rules demonstrate moderate confidence, that is, the strength of the association between the antecedents and the consequents. Another way to think of confidence is how much of an overlap there is in the set of transactions containing the antecedent and the set of transactions containing the consequent. The greater the overlap, the higher the confidence. (Zhang, n.d.)

The rules also demonstrate strong lift. Lift is the ratio of confidence divided by expected confidence. In other words, lift measures how likely the antecedent and consequent are to occur together in a transaction compared to the combined probabilities that each will appear separately. Values greater than 1.0 show strong relationships. (Chauhan, 2020)

# D – Analysis

As mentioned above, the generated association rules show evidence of a strong likelihood of appearing together in pharmacy transactions. This analysis is based on actual transaction data, so it is eminently practical rather than theoretical. Inventory managers should evaluate the study results and plan to rearrange the positioning of medications in the pharmacy to place often-co-ordered drugs together. Further, IT managers may need to update order processing software to recognize when two or more medications in the same order are co-located to direct pharmacy employees to gather both drugs at once rather than iteratively.

This graph shows the top ten medications most frequently associated with others in the list of association rules we generated.

A black screen with many small colored text

Description automatically generated with medium confidence

A graph of a bar chart

Description automatically generated

This list would be an excellent starting point for planning new placements within the pharmacy.

# E – Panopto Presentation

A demonstration of this study is available [here](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=de06b34a-88a2-4cd4-b574-b1fd0154e6e5).

# F – Third-Party Code References

Herfert, F. (2019, 07 18). *Extract string from rules frozensets*. Retrieved from StackOverflow: https://stackoverflow.com/questions/52291739/extract-string-from-rules-frozensets

Kamara, K. (2024, 09 09). *Market Basket Analysis*. Retrieved from https://srm--c.vf.force.com/servlet/fileField?retURL=https%3A%2F%2Fsrm--c.vf.force.com%2Fapex%2FCourseArticle%3Fid%3DkA0S60000000ilpKAA%26groupId%3D%26searchTerm%3D%26courseCode%3DD212%26rtn%3D%252Fapex%252FCommonsExpandedSearch&entityId=ka0S60000004FBlIA

Yasik, I. (2023, 10 04). *Market Basket Analysis & Apriori Algorithm using Zhang’s Metric*. Retrieved 09 30, 2024, from Medium: https://medium.com/@iuyasik/market-basket-analysis-apriori-algorithm-using-zhangs-metric-708406fc5dfc

# G – Referenced Works

Chauhan, A. (2020, 04 26). *Association Rule Mining — concept and implementation*. Retrieved 10 02, 2024, from Medium: https://medium.com/analytics-vidhya/association-rule-mining-concept-and-implementation-28443d16f611

Noble, J. (2024, 06 09). *What is the Apriori algorithm?* Retrieved 10 01, 2024, from IBM Topics: https://www.ibm.com/topics/apriori-algorithm

Pai, A. (2024, 07 05). *Market Basket Analysis: A Comprehensive Guide for Businesses*. Retrieved 10 01, 2024, from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-market-basket-analysis

Zhang, L. (n.d.). *Understanding Support, Confidence, Lift for Market Basket (Affinity) Analysis*. Retrieved 10 02, 2024, from The Data School: https://www.thedataschool.co.uk/liu-zhang/understanding-lift-for-market-basket-analysis