Market Basket Analysis of Telcom Customer Churn Data

Mike Mattinson

Western Governors University

D212: Data Mining II

Task 3: Association Rules and Lift Analysis

Dr. Kesselly Kamara

May 17, 2022

Revision 1

Abstract

Telecom customer data will be analyzed for feature selection using principal component analysis (PCA). The dataset consists of 50 features associated with 10,000 customer records. The purpose of this analysis is reduce number of features by finding and removing irrelevant data. The analysis will reduce the number of principal components to 39. While using Kaiser criterion, the analysis will reduce the number of principal components to 26.

Keywords: Telecom. Churn. Data Mining. Market Basket Analysis. Association Rules and Lift Analysis.

Scenario 1

One of the most critical factors in customer relationship management that directly affects a company’s long-term profitability is understanding its customers. When a company can better understand its customer characteristics, it is better able to target products and marketing campaigns for customers, resulting in better profits for the company in the long term.

You are an analyst for a telecommunications company that wants to better understand the characteristics of its customers. You have been asked to perform a market basket analysis to analyze customer data to identify key associations of your customer purchases, ultimately allowing better business and strategic decision-making.

List of Tables

[Table 1 *Description of the prepared data prior to scaling* 7](#_Toc103237067)

List of Figures

[Figure 1 *D1. Determine matrix of all the principal components* 9](#_Toc103237070)

# Part I: Research Question

A. Describe the purpose of this data mining report by doing the following:

## A1. Propose one question relevant to a real-world organizational situation that you will answer using market basket analysis.

## A2. Define one goal of the data analysis. Ensure that your goal is reasonable within the scope of the scenario and is represented in the available data.

# Part II: Market Basket Justification

B. Explain the reasons for using market basket analysis by doing the following:

## B1. Explain how market basket analyzes the selected dataset. Include expected outcomes.

## B2. Provide one example of transactions in the dataset.

## B3. Summarize one assumption of market basket analysis.

# Part III: Data Preparation and Analysis

C. Prepare and perform market basket analysis by doing the following:

## C1. Transform the dataset to make it suitable for market basket analysis. Include a copy of the cleaned dataset.

## C2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate the error-free functionality of the code.

## C3. Provide values for the support, lift, and confidence of the association rules table.

## C4. Identify the top three rules generated by the Apriori algorithm. Include a screenshot of the top rules along with their summaries.

Table   
Description of the prepared data prior to scaling

Here is the code:

# Part IV: Data Summary and Implication

D. Summarize your data analysis by doing the following:

## D1. Summarize the significance of support, lift, and confidence from the results of the analysis.

## D2. Discuss the practical significance of the findings from the analysis.

## D3. Recommend a course of action for the real-world organizational situation from part A1 based on your results from part D1.

Figure   
D1. Determine matrix of all the principal components

Notes. Figure 1 is the correlation matrix for all numerical data, including the newly created categorical dummy features. To help identify the features with high correlation, the names of the features with correlation above 0.7 are listed in the figure notes.

# Part V: Attachments

E. Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.

Note: The audiovisual recording should feature you visibly presenting the material (i.e., not in voiceover or embedded video) and should simultaneously capture both you and your multimedia presentation.

Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link provided below. To access Panopto's website, navigate to the web link titled "Panopto Access," and then choose to log in using the “WGU” option. If prompted, log in using your WGU student portal credentials, and then it will forward you to Panopto’s website.

To submit your recording, upload it to the Panopto drop box titled “Data Mining II – OFM3.” Once the recording has been uploaded and processed in Panopto's system, retrieve the URL of the recording from Panopto and copy and paste it into the Links option. Upload the remaining task requirements using the Attachments option.

F. Record all web sources used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.

G. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

H. Demonstrate professional communication in the content and presentation of your submission.

References

Albon, C. (2017). Drop Highly Correlated Features. Retrieved from https://chrisalbon.com/code/machine\_learning/feature\_selection/drop\_highly\_correlated\_features/

Alkhayrat, M., Aljnidi, M., & Aljoumaa, K. (2020). A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. Retrieved from https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-0286-0

Arvai, K. (2022). K-Means Clustering in Python: A Practical Guide. Retrieved from https://realpython.com/k-means-clustering-python/#: :text=The SSE is defined as,try to minimize this value.&text=The purpose of this figure,centroids is an important step.

Bex, T. (2021, April 13). *How to Use Pairwise Correlation For Robust Feature Selection*. Retrieved May 8, 2022, from How to Use Pairwise Correlation For Robust Feature Selection: https://towardsdatascience.com/how-to-use-pairwise-correlation-for-robust-feature-selection-20a60ef7d10

Boyle, T. (2019, March 25). *Feature Selection and Dimensionality Reduction*. Retrieved May 8, 2022, from Feature Selection and Dimensionality Reduction: https://towardsdatascience.com/feature-selection-and-dimensionality-reduction-f488d1a035de

Brownlee, J. (2019). How to Calculate Principal Component Analysis (PCA) from Scratch in Python. Retrieved from https://machinelearningmastery.com/calculate-principal-component-analysis-scratch-python/

Bruce, P. C., Gedeck, P., Shmueli, G., & Patel, N. R. (2019). *Data Mining for Business Analytics Concepts, Techniques and Applications in Python.* Wiley & Sons, Incorporated, John.

Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical Statistics for Data Scientists 50+ Essential Concepts Using R and Python.* O'Reilly Media, Incorporated.

Fenner, M. (2018). *Machine Learning with Python for Everyone.* Addison Wesley.

Galarnyk, M. (2017). PCA using Python (scikit-learn). Retrieved from https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems.* O'Reilly Media.

h1ros. (2019). Drop Highly Correlated Features. Retrieved from https://h1ros.github.io/posts/drop-highly-correlated-features/

Kaloyanova, E. (2020, March 10). *How to Combine PCA and K-means Clustering in Python?* Retrieved May 8, 2022, from How to Combine PCA and K-means Clustering in Python?: https://365datascience.com/tutorials/python-tutorials/pca-k-means/

Larose, C. D., & Larose, D. T. (2019). *Data Science.* Wiley.

Parveez, S., & Iriondo, R. (2018). Principal Component Analysis (PCA) with Python Examples — Tutorial. Retrieved from https://pub.towardsai.net/principal-component-analysis-pca-with-python-examples-tutorial-67a917bae9aa

ProjectPro. (2022). How to drop out highly correlated features in Python? Retrieved from https://www.projectpro.io/recipes/drop-out-highly-correlated-features-in-python

Sawlani, D. (2017). Customer Churn Prediction Model. Retrieved from https://rstudio-pubs-static.s3.amazonaws.com/344958\_a3c32d38a83043c8ae8fc13cd6abbda4.html

Sharma, A. (2021). PCA or Principal Component Analysis on Customer Churn Data. Retrieved from https://medium.com/data-science-on-customer-churn-data/pca-or-principal-component-analysis-on-customer-churn-data-d18ca60397ed

Srinivas. (2022). Reducing Telecom Churn with PCA and Modeling. Retrieved from https://www.kaggle.com/code/manoharsrinivas/reducing-telecom-churn-with-pca-and-modeling/notebook

VanderPlas, J. (2016). *In Depth: Principal Component Analysis.* O'Reilly. Retrieved from https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html