Unsupervised Learning Algorithms for Customer Segmentation of Financial Firms: The case of 3Rivers Federal Credit Union

Aman Bajpayee [1] Danielle Blackford [1] Monika Vyas [1] Rishikesh Fulari [1] Rohan Patel [1] Yashi Yadav [1]
Supervisors: John Dever [2] Brent Hepner [2] Alessandro Maria Selvitella [1]

[1]Purdue University Fort Wayne [2]3Rivers Federal Credit Union

Company overview

Since 1935, 3Rivers Federal Credit Union has been empowering our community to achieve financial wellness by offering personalized service, tools, and education. The member-owned, not-for-profit cooperative has $2.2 billion in assets, more than 110,000 members, 23 branches, and nearly 500 employees. 3Rivers offers a wide range of financial solutions, in addition to trustworthy, lifelong guidance and relationships. For more information, visit `3riversfcu.org`.

Project description

This project focuses on segmenting 3Rivers customers into groups based on similar needs and behaviors. These clusters will be used to study 3Rivers member data and tailor advertising and marketing campaigns to meet customer needs and enhance their experiences. This purpose requires unsupervised learning algorithms as there will be no segment labels provided for any of the members.

Desired outcome/Question to be answered

Given a set of demographic and past transaction data, a 3Rivers member is assigned to a particular segment and/or for each segment a "score" is generated which measures how well the member aligns with that segment. We want to determine the optimal number of segments, as well.

Data Sources and Tools

A series of flat file de-identified datasets have been provided to PFW students by 3Rivers Federal Credit Union. These files include variables such as accounts held, balances, transaction information, and some demographic data. These files have been provided in RDS and HDF5 formats. Data dictionaries have been also provided with each file.

Models have been built using either R or Python.

Research Objectives

- **Objective 1:** Leverage historical data to segment customers into different clusters using unsupervised clustering algorithms.
- **Objective 2:** Study and conduct elaborate research to derive insights about customer behavior, which would in turn help 3Rivers serve their customers better.
- **Objective 3:** Evaluate the risk associated with customer accounts and financial transactions. Develop a comprehensive risk strategy that enables the bank to proactively manage and mitigate potential financial risks.

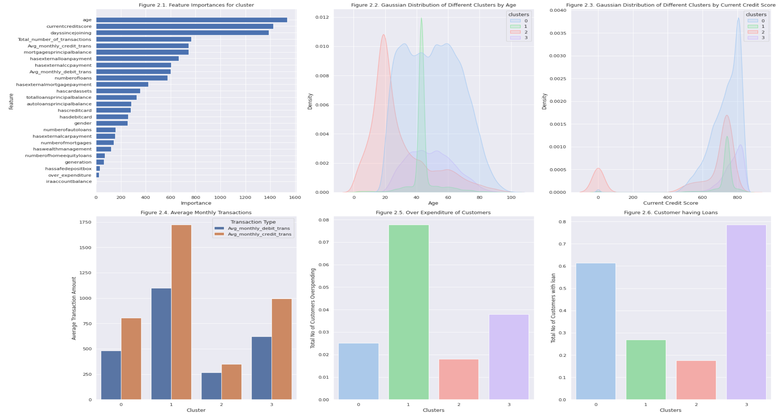
Study Methodology

The original dataset consisted of eight tables, each detailing customer information related to member's accounts, individual demographics, and lending and transaction history. We filtered the data to focus solely on a predetermined sub-population, which only includes adult customers with active, personal accounts, aged at least three months.

Upon restricting data to the appropriate population and removing outliers and leverage values, we each perform clustering on our designated tables. Methods of clustering include K-Means++, K-Means and K-Prototype assuming either four or eight clusters. The most informative, conclusive results are discussed in greater detail in the following sections.

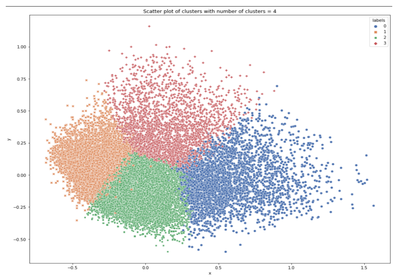
Results and Discussion

We have segmented customers into four distinct groups based on historical customer behaviour data and demographics.



Observations:

- **Cluster 0:** Age range (over 20): Despite having an average income and a high percentage of loan-holding customers, **62%**, their spending habits are balanced, with no tendency to overspend.
- **Cluster 1:** Age range (mid 40s): Consists mainly of individuals with the highest income levels among all clusters. Although only 27% have loans, 8% tends to spend more than they earn.
- **Cluster 2:** Age range (10 to 30): Lower incomes and prudent spending habits, focusing on building their credit scores.
- **Cluster 3:** Age range (20 to 60): Has good income and excellent credit scores, over 800, coupled with healthy spending habits. Approximately 1.7% in this group have loans, but their financial behavior suggests they manage their debts effectively.



Observations:
We also grouped the data based on census information, which is comprised of individual's income range based on different regions, demographics, education, employment status, race and etc. The graph above represents the four clusters we obtained using the same data.

Conclusions

- **Clustering based on user information:** Customers can be grouped into four distinct groups based on transaction data and user information.
- **Clustering based on census data:** Overall customers can also be grouped into four groups based on the census information.

What is already known about this subject?

- **Prevalence of Traditional Segmentation Methods:** Financial institutions have traditionally segmented customers based on basic demographic data, geographical location, or simple transaction history, which provided a broad but often superficial understanding of customer groups.
- **Challenges in Traditional Segmentation:** Conventional methods often fail to capture the complexities of customer behaviors and changing preferences, leading to less effective marketing and customer service strategies.
- **Emergence of Data-Driven Approaches:** There is a growing trend in the finance industry towards adopting more sophisticated, data-driven methods, including machine learning and AI, for dynamic and accurate customer segmentation.

What does this study add?

- **Innovative Use of Unsupervised Learning:** This study leverages advanced unsupervised learning algorithms for customer segmentation, offering a novel approach in the financial sector to understand and categorize customer behaviors in a more nuanced manner.
- **Advanced Customer Insight:** The application of these algorithms provides deeper insights into customer preferences and behavior patterns, enabling the development of more personalized and effective marketing and service strategies.
- **Model Adaptability and Scalability:** The models presented in this study are adaptable and scalable, capable of handling diverse datasets and evolving to meet the changing needs of the financial institution, thereby offering a sustainable and versatile tool for customer analysis. These can be helpful in providing more targeted services to the customer base.

Acknoweldgements

This project is part of the Indiana Data Mine initiative at Purdue University Fort Wayne. The work has been performed by the students during Fall 2023 as part of their class work for MA 598 Corporate Partners I, sponsored by 3Rivers Federal Credit Union and instructed by Prof. Selvitella. Prof. Selvitella acknowledges the support of Lilly Endowment Inc. through the "Indiana Data Mine" grant. We thank 3Rivers for the generous support and John Dever and Brent Hepner for the kind guidance and contributions.

Lilly Endowment Inc.
A private foundation since 1937