

REPORT

Big Data and Credit Unions: Machine Learning in Member Transactions

Philipp Kallerhoff, PhD, Jumiya

ACKNOWLEDGMENTS

I would like to express my greatest gratitude to Mark Meyer, George Hofheimer, Ben Rogers, and Brent Dixon for shaping this report. The environment and support they set up at the Filene Research Institute was highly inspiring, and my numerous discussions with them were essential sources of motivation and triggers for new ideas. Furthermore, I would like to thank the participants in this study. The focused discussions, paired with openness, kindness, and hospitality, were unique. This report owes a lot to my interaction with them.

Table of Contents

4	EXECUTIVE SUMMARY
6	CHAPTER 1 Introduction
8	CHAPTER 2 Project Beginnings
10	CHAPTER 3 Methods and Results
16	CHAPTER 4 How to Use Machine Learning
19	CHAPTER 5 Conclusion
22	APPENDIX
23	ENDNOTES
24	LIST OF FIGURES
26	ABOUT THE AUTHOR
27	ABOUT FILENE

Executive Summary

Overview

Credit unions can take a cue from big corporations by using transactional data to mine for useful insights around key business questions. Big data used right can help improve underwriting, predict members' next products, and help members build wealth.

ABOUT THE AUTHOR



Philipp Kallerhoff, PhD
Jumiya

by Ben Rogers
Research Director

Facebook reports that its users have piled more than 100 petabytes of data into its social maw. To give you a sense of that dizzying scale, that's the equivalent of 102,400 fancy 1-terabyte desktop storage drives. If you lined up 100,000 of those 3-inch drives, you'd have to walk about five miles to reach the end. That's a lot of baby photos. Get walking.

Facebook joins big companies like FedEx, Amazon, Walmart, Salesforce, and IBM in the new game of "big data"—searching for interesting connections in a sea of bytes that no human or team of humans could ever possibly synthesize. But advances in cloud storage, computing power, and analytics mean that even modestly sized credit unions should consider how they can use the trove of transactional data that run through their systems every year.

What Is the Research About?

For this research, five credit unions in the United States and Canada proffered their members' anonymous profile information and transaction details to the researcher, who used variables as diverse as gender, product balances, credit score, income, and transaction amounts to search for revealing correlations.

The credit unions were each looking for different things, and the plasticity of big data means that, with the right tools and the right inputs, you can discover very different things. Several credit unions wanted to better understand how members cycle through different products at different stages of their membership, all the better to introduce the right products at the right time. Another was interested in using transactional data to improve underwriting, searching for insights that would fuel more originations without pushing up delinquencies. And finally, one credit union wanted to see how to simultaneously promote wealth and profitability among its members.

What Are the Credit Union Implications?

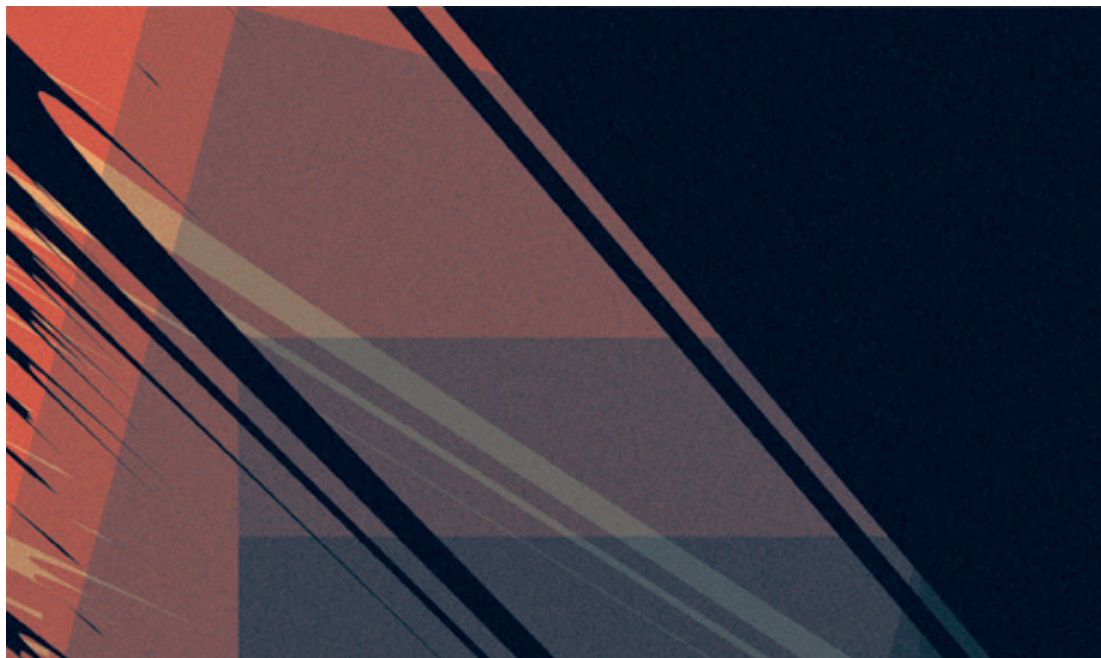
Machine learning is a branch of artificial intelligence that focuses on the construction and study of systems that can learn from data. Because a machine learning project is only as helpful as the data that flow into it, the

participating credit unions got the most specific insights. But their combined data still offer generalizable findings for all credit unions. Among them:

- ➔ **Product progression.** What and when people buy and how much debt they're carrying can indicate how likely they are to upgrade (or close) their accounts. The cluster analysis undertaken here can predict the next best product with 30% accuracy. The best predictors are the balance-to-income ratio, the expense-to-income ratio, and the balance of loans to savings. A dynamic system that tracked these changing ratios at the member level could generate automatic marketing messages when members reached the likely-to-switch threshold.
- ➔ **Improved credit scoring.** Transactional data add another dimension to traditional credit scores. This research shows that transaction amounts and the number of transactions are positively correlated with creditworthiness. Used correctly, these insights align perfectly with credit union values, allowing lenders to use transactional information to take different risks on members than a standard credit score would allow.
- ➔ **Cleaner data needed.** A prerequisite for developing these and other models is a well-maintained database with as much transactional detail as possible. The credit unions that can capture transaction types and locations will come out ahead, because transaction origin correlates highly with credit scores and helps to predict future financial products.

Simply having a 100-petabyte cache like Facebook or 100 gigabytes like a typical credit union doesn't guarantee any insights. Credit unions are best served when they start with a goal and only then decide whether machine learning can get them there. But don't sit this one out. Big data is here to stay.

Big Data and Credit Unions: Machine Learning in Member Transactions



CHAPTER 1

Introduction

Companies as varied as Amazon, Google, Walmart, and Wells Fargo are turning to “big data” for customer insights that will help them serve clients and capture market share. Big data is the analysis of huge data sets. Individual credit unions may not have the resources of a corporate giant, but advances in data storage and software tools mean that credit unions can start using similar tools and deriving similar value. Searching for insights into member life cycles, improved underwriting, and profitability cues, we applied big data and machine learning to millions of transactional data points from five credit unions.

The data were combined across the five participating credit unions—three in Canada and two in the United States—which provided 250 million transactions for analysis. The findings show that some simple patterns evolve using machine learning and big data. In

particular, we found that members follow simple paths during their life cycle and adopt different consumer products at each stage.

Machine learning is a branch of information technology that is primarily concerned with the construction and study of systems that can learn from data. The core of machine learning deals with representation and generalization. Representation in this study considers the single data instances of the transactions of each member. Generalization then uses a system that performs well on unseen data instances and predicts member behavior.

Machine learning is powerful, but it is not easy. Challenges include data storage as well as visualization. It's one thing for a machine to generate clusters and draw correlations; it's another thing for those data to be useful to managers. The analysis in this report is based on a data set from 500,000 credit union members and 250 million transactions over a five-year time frame, reported in a 10-dimensional space.

It's one thing for a machine to generate clusters and draw correlations; it's another thing for those data to be useful to managers.

To build such a system based on machine learning and big data, we assume that past behavior is a good predictor of future behavior and that transactional data in consumer finance provide a new level of detail for these analyses.¹ In fact, we found patterns predicting member behavior based on their transactions and external factors such as the state of the economy. Questions that drove the research included: What are the patterns that members mature into during their life cycle at a credit union? How do these patterns predict the next best product for each member? And how do past transactions predict future delinquencies?

The analysis in this report is based on a data set from 500,000 credit union members and 250 million transactions over a five-year time frame, reported in a 10-dimensional space.

To analyze patterns that members mature into during their life cycle, we grouped them into clusters and then noted which clusters they switched to later on. A cluster is a group of members who are more similar to each other than to those in other clusters. The discriminator of each cluster is the product mix, which consists of the products that all members of a cluster have in common and that are most different from the products held by members in the other clusters.

In this analysis we describe patterns in how members switched from one cluster to another. A switch is notable because it predicts which products might interest members in the near future. Transactional data and the corresponding detailed information about each member increased the prediction level of these switches.

An example of this method is Amazon.com generating recommendations to its shoppers based on what similar users bought or looked at. Here we report an algorithm that recommends additional products to a credit union member. The analysis stretches from the next best product to different profitabilities for the credit union and the potentially changing wealth of the member in the future.

Other examples of machine learning and big data are seen in the hedge fund and investment banking industries, where analysts collect large data sets and build models that analyze financial market behavior. For these models, input variables are selected as market prices (e.g., the S&P 500) as well as fundamental data from balance sheets and financial reporting and even macroeconomic data (e.g., GDP performance or inflation). These variables are then integrated and used as an input vector to a large-scale model and applied to financial markets. In this research study we take a similar approach, first selecting input variables and then predicting the behavior of credit union members.

CHAPTER 2

Project Beginnings

The impetus for this credit union project was the Jumiya project (www.jumiya.com). The Jumiya project was developed in 2012 during Singularity University's Graduate Studies Program, which is supported by Google, NASA, and others in Mountain View, California. The program seeks to build projects that will positively impact a billion people over the next 10 years. During the summer 2012 session, the Jumiya project was supported by the National Credit Union Roundtable and advised by Peter Kellner, founding partner at Richmond Global and cofounder of Endeavor, a global high-impact entrepreneurship investment fund.

Jumiya is now being developed as a wellness platform that rewards active and healthy lifestyles with access to financial loans and better interest rates, as well as real-world rewards, in cooperation with banks and credit unions. The Jumiya team scientifically demonstrated that individuals who exercise regularly, eat well, and rarely drink or smoke are more likely

to pay their bills on time and save monthly. This is great news not only for active individuals but also for financial providers, which are happy to reward such behavior.

Those signing up for a Jumiya-supported account will provide access to the stream of data from their self-tracking devices, such as the Nike FuelBand. Jumiya is particularly interested in physical activity levels, but it can also include models for sleep and diet. These data will be aggregated and displayed through a master dashboard, making it easy for the user to visualize and understand the data. The use of predictive models and gamification will motivate further physical improvement and achievement of goals. As users' activity levels increase and their goals are achieved, they may redeem points for better rates or loans or for discounts at local businesses.

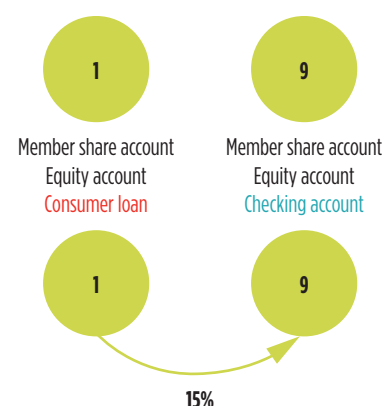
Building on insights gleaned from the Jumiya project, we wanted to construct a prediction model for member behavior using transactional variables and derived measures such as expense-to-income ratio and the balance of loans. Additionally, we used the S&P 500 index as an external variable in light of the insight from the Jumiya project that people do not live in a bubble; they are affected financially by external factors and by their own behavior in very different areas of life.

In this research study the first step was to analyze how members improve and mature during their life cycle at the credit union. For this analysis, members were clustered in an optimal number of groups with the highest out-of-sample validation at several points in time. That is, we calculated the likelihood that each member would be associated with one of the clusters. As members evolve during their life cycle at the credit union, they are associated with different clusters, each of which represents a different product mix. In addition, a member's association with one cluster predicts which cluster that member will be associated with in the future. This type of analysis is good at projecting a member's product path; it allowed us to predict a member's next likely product, such as a credit or loan product, term, or demand product.²

The initial goal of this research study was to identify clusters of members and determine the likelihood for a member to switch between those clusters. Figure 1 shows the products of two example clusters, 1 and 9, and the likelihood that the members of cluster 1 will switch to cluster 9. Members in cluster 1 have in common a member share account, equity account, and consumer loan product. The figure shows that members in cluster 1 tend to replace the consumer loan product with a checking account. The overall probability of migration between the two clusters is 15%, but this might differ for a single member. Here we used the transactional data to provide an indicator that predicts shifts between groups with a likelihood above 30%.

FIGURE 1

NEXT BEST PRODUCT ANALYSIS



The initial goal of this research study was to identify clusters of members and determine the likelihood for a member to switch between those clusters.

Past transactions also provide insight into a member's solvency and are therefore good predictors of a member's likelihood to become delinquent or even default on loans. Again, we applied machine learning to the vast amounts of transactional data and developed a simple algorithm to improve the prediction accuracy of credit risk.

Therefore, we processed transactional data and selected features to predict the likelihood of default or delinquency. Using machine learning defaults, delinquencies could be predicted with more than 40% accuracy.³

CHAPTER 3

Methods and Results

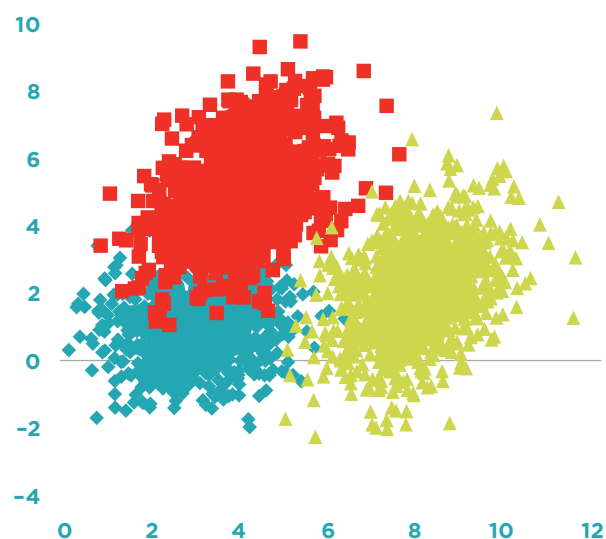
Cluster Analysis

Members were clustered according to their product mix. Clustering is the grouping of a set of data points in such a way that data points in the same cluster are more similar to each other than to those in other clusters. Clustering is part of exploratory data mining and a common technique for statistical data analysis.

There are various algorithms for clustering that differ in their notion of what constitutes a cluster and how to efficiently find clusters. For this analysis, we used a popular mechanism called k -means clustering. This mechanism partitions the data points into k clusters. Each data point represents a single member and his or her current products—specifically, a vector of his or her account balance for all possible products. If the member did not have a particular product, the account balance was set at zero. Figure 2 shows an example of k -means clustering with three

FIGURE 2

EXAMPLE OF k -MEANS CLUSTERING



different clusters. Dots are grouped without any knowledge about the cluster affiliation and only with the selection of the number of different clusters (in this case, three).

The optimal number of clusters is selected by an out-of-sample test. Figure 3 plots the centers for each of 10 clusters against the products of one of the participating credit unions. Only the top 10 products are displayed. The colors indicate the likelihood that the members of a cluster are using a product. For example, for this credit union the algorithm found a cluster (example cluster 1) whose members are very likely to have the member share account, nonredeemable member equity account, and consumer term loan products, with likelihood higher than 70%.⁴ Since the products are sorted according to their number of occurrences in the database, the first product listed is very likely in most of the clusters.

However, the clusters differ widely and represent quite different member groups. Some of the clusters include members who only use a very small set of products, while other clusters engage in almost all products.

Another key result from this analysis is that there is always one cluster consisting of members who have no products, such as cluster 3 at the credit union in this example. Detailed analysis showed that the members of this group had actually left the credit union and that their products were mainly inactive. This is an important cluster type to study: when members switch into it, they very likely leave the credit union altogether. This cluster is also relevant when analyzing members who are new to the credit union, since they do not have any products prior to joining.

Life Cycle Analysis

In general, members will switch clusters at least once during their life cycle at a credit union, and each member is associated with a maximum likelihood of being in a certain cluster at any single point in time. That likelihood is calculated as the distance between the products the member has and the center of each cluster. Each member is represented as a vector of his or her products and the balance of his or her products in the associated fields. The cluster centers are taken from the prior analysis.

FIGURE 3

RESULTS OF THE CLUSTER ANALYSIS

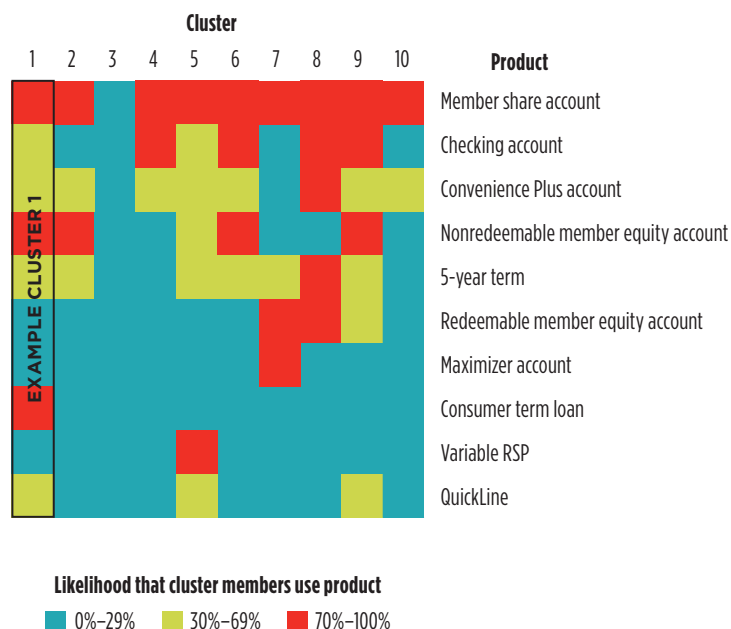


FIGURE 4

DISTANCE MEASURE

$$D = \left(\sum_{i=1}^n |x_i - y_i|^2 \right)^{1/2}$$

Distance (D) between vector of actual products (x) of member (i) and centers of each cluster (y) in the n dimensional space of the top products

FIGURE 5

INPUT VARIABLES

Original data	Derived measures	External variable
1. Gender (1 = male, 2 = female)	1. Balance-to-income ratio	1. S&P 500 price on date of transaction
2. Birth date	2. Expense-to-income ratio	
3. Marital status (1 = single, 2 = married)	3. Income change (month over month)	
4. Type of customer	4. Number of transactions	
5. Customer status	5. Number of loans	
6. Type of product	6. Balance of loans	
7. Product balance	7. Balance of loans to savings	
8. Type of transaction	8. Total inflow	
9. Amount of transaction	9. Total outflow	
10. Credit score		

Shifts between clusters can be predicted by looking at a member's past behavior and product changes in the member base. To predict a member's switch to a different cluster, we used a statistical analysis called support vector machine (SVM) and trained an algorithm to the member's past behaviors. The input variables to the SVM are based on the original transactional data, derived measures, and an external variable.

The original data were provided in the data sets for each transaction. A negative transaction amount is considered an expense, whereas a positive amount is considered income.

The derived measures were based on the original data and were calculated for each transaction date. For example, the balance-to-income ratio was calculated as the balance at the time the transaction occurred divided by the income in the prior month.

One of the biggest insights from the Jumiya project is that people don't live in a bubble; economic data sets, health, transportation, and energy consumption all affect members' profitability and wealth. The Jumiya project particularly analyzes the relationship between health behavior and financial wealth. The main problem that was identified is that there is a vicious cycle of poverty connecting healthcare cost and lack of financial access. In the United States, 62% of bankruptcies are due to medical debt, and this percentage is rising steadily.⁵ Most medical debt is related to treatment of obesity, diabetes, and heart disease,

which are preventable with exercise, change of diet, and a generally more active lifestyle. To summarize, health is closely linked to personal financial performance and ability.

As a proxy of external factors beyond the transactional data, we included the S&P 500 price on the date of the transaction, which provides some insight into members' behavior. During the economic crisis of 2008, many people lost their jobs and defaulted on loans. Also, the crisis changed the economic outlook, and people had to plan financially for those times. External factors such as this provide an additional level of prediction for member behavior.

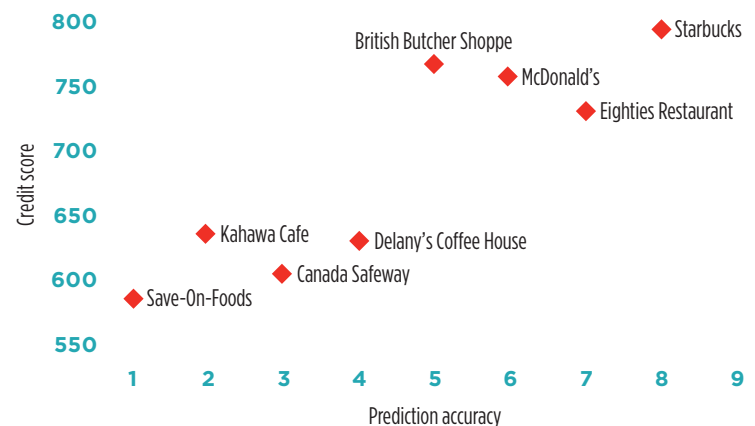
FIGURE 6
S&P 500 PRICE, 2003–2013



Prediction of the Member's Credit Score

Simple member behavior, such as which restaurants, grocery stores, and coffee shops members frequent, can predict their credit score. Figure 7 shows different credit score predictions for different stores. For example, members who frequent the British Butcher Shoppe very likely have a credit score of around 800. Members who shop at Save-On-Foods, on the other hand, very likely have a credit score of around 600. The prediction accuracy is given as a score, with 1 representing the highest prediction accuracy for a certain store and 20 the lowest prediction accuracy in an out-of-sample test.

FIGURE 7
STORE CHOICE AS PREDICTOR OF MEMBER CREDIT SCORE



Prediction of Cluster Switches

The input variables were calculated for each transaction. The first step in calculating the probability of a member switching to a different cluster was to select only certain input features. The features were selected with an analysis of variance approach. Here the observed

variance of moving to a different cluster was partitioned into components attributable to the different input factor variations. In particular, we rejected input variables with an explained variance of less than 2%.

The resulting input variables were then used to calculate a member's probability of switching to another cluster using an SVM applied in a regression analysis. That is, it returned not only the category label for a new data point, but also the distance to the next data point. The advantage is that the SVM returns not only the category label but also the likelihood that the data point is associated with that category.

Figure 8 shows the influence of several input variables on the model across the participating credit unions. Some of the variables were excluded from the analysis during feature selection and have zero weight. The values indicate the impact of each variable on the prediction algorithm, with absolute weights indicating that the variable is highly related to switching to another cluster. The sign of each weight normalizes the values of the different input variables.

Note that the input variables are transformed into numeric values, as described in Figure 5. For example, the gender input becomes 1 if the member is male and 2 if the member is female. The same logic is applied to marital status, such that a single member is denoted with 1 and a married member with 2. Calculating the weighted average across these variables and comparing it to an offset value emphasizes the importance of changes in these variables over time. If a member gets married, that input variable changes and the member receives a higher score. As a result, the member is more likely to switch clusters.

The highest impact on a member's switch to a different cluster is seen in the credit score, the balance-to-income and expense-to-income ratios, the balance of loans to savings, and the S&P 500. These variables also provide insight into the current status of the member.

A member's association with one cluster predicts his or her likelihood of switching to one of the other clusters. Figure 9 shows the

FIGURE 8

MOST LIKELY TO SWITCH

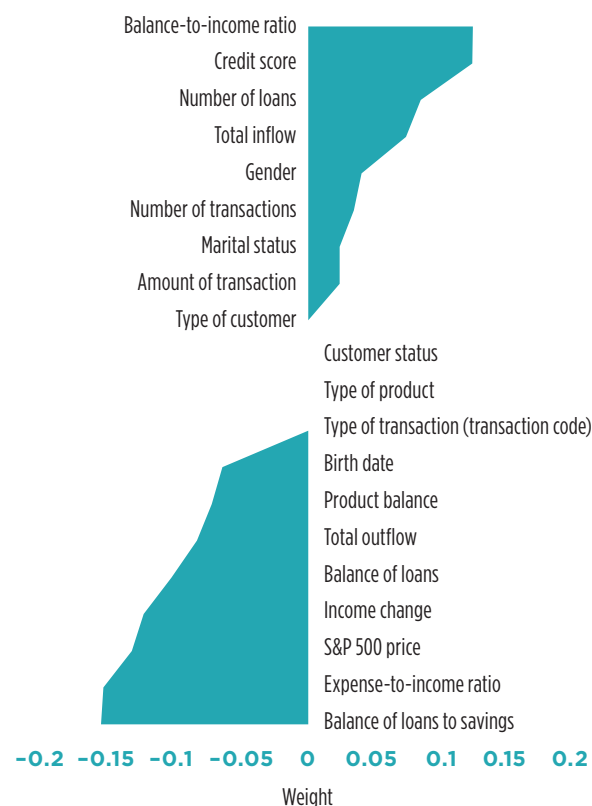
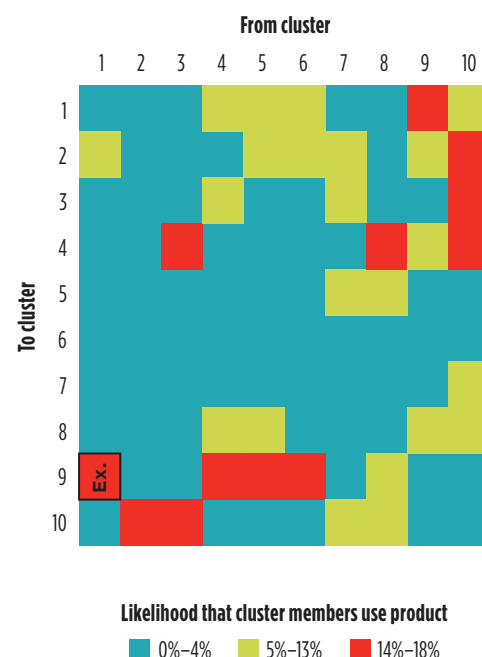


FIGURE 9

NEXT BEST CLUSTER ANALYSIS



likelihood of the next cluster (y-axis) for each of the calculated clusters (x-axis). Note that the SVM is only applied to predict the cluster switch with the highest likelihood.

The likelihood that a member will switch to another cluster, indicated by the colors in the figure, is calculated based on his or her past behavior. For this analysis, members were grouped into clusters each quarter. The likelihood of members switching to the next cluster was then calculated as the probability of them being in the new cluster given that they were in the old cluster in the previous quarter.

For example, members in cluster 1 are very likely to switch to cluster 9 during the next quarter. Referring back to Figure 3, members in cluster 1 generally have a member share account, a nonredeemable member equity account, and a consumer term loan. If they switch to cluster 9, they generally replace the consumer term loan with a checking account. Generally the maximum likelihood of switching to the next cluster across clusters and credit unions is around 20%.

The likelihood of staying in the same cluster as in the previous quarter is always the highest for all credit unions and clusters. In Figure 9 this likelihood was set to zero, but it can easily be calculated as one minus the sum of all likelihoods to switch to a different cluster. Also, in all credit unions we found at least one cluster whose members were very unlikely to switch to a different one.

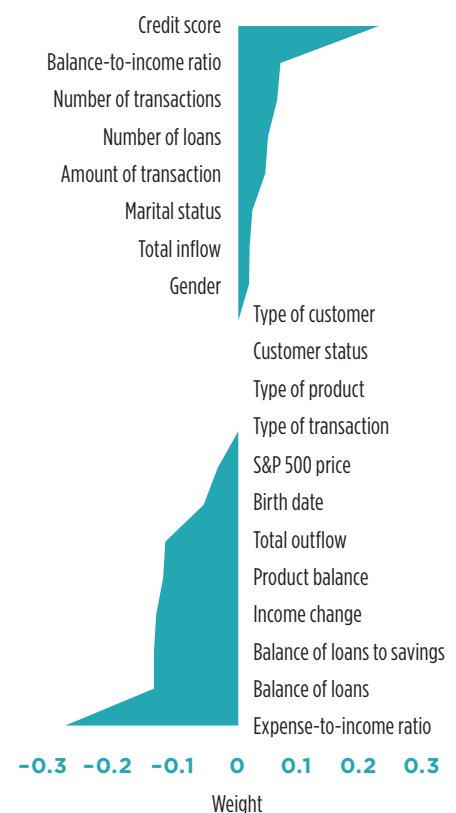
Prediction of the Member's Credit Risk

We also used the input variables to predict a member's risk of defaulting or becoming delinquent on one or more products. Again, we applied feature selection and the machine learning algorithm; features with less than 2% explained variance of the credit score were not considered for input to the model.

Here the labels for the SVM were not switches to another cluster but a label of +1 if the member became delinquent and -1 if he or she did not. The SVM was then tested with new and unseen data points.

Figure 10 shows the values of normal vector w , which is weighting the input variables and works as a kind of filter, as an average across the participating credit unions. The original credit score has the highest impact, which underlines the basis of this original value as a starting point. Beyond that, the expense-to-income ratio has a high impact on the new credit scoring as well as the balance of loans and the ratio of balance of loans to income.

FIGURE 10
INPUT FILTER



The expense-to-income ratio has a high impact on the new credit scoring as well as the balance of loans and the ratio of balance of loans to income.

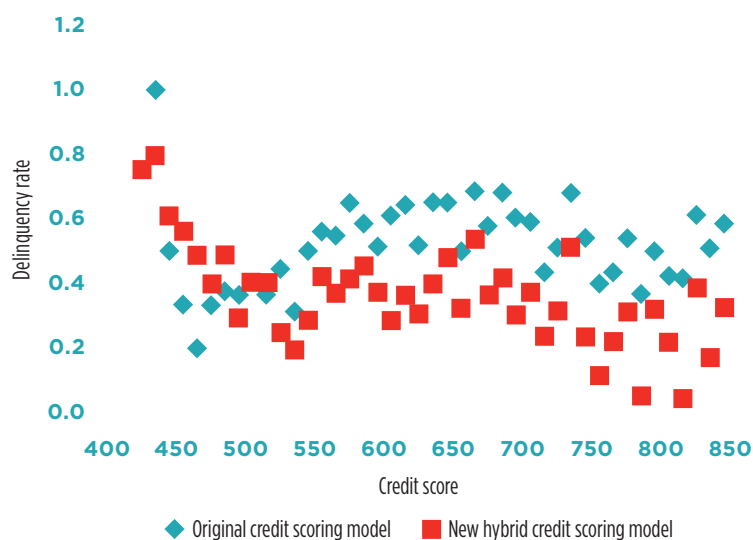
This method predicted delinquencies with 40% accuracy on an out-of-sample test. The algorithm also improved on the traditional credit score by 10%, as shown in Figure 11. This calculation of improvement is based on a comparison between the number of misclassifications in the original and improved scoring methodologies.

The figure also shows the difference in delinquency rates between the two scoring models for different credit scores. The blue diamonds are average delinquency rates for different credit scores in the original data. The red squares are the average delinquency rates resulting from the new hybrid model, which is based on the original credit score and the additional input variables from this study.

The difference in delinquency rates is particularly high for high credit scores. Using the original credit scoring model, members with high credit scores become delinquent almost 50% of the time. With the hybrid scoring model, that number drops to less than 20% for credit scores above 800. Note that the overall high delinquency rates result from a very limited database and that the label was associated with delinquency over the member's entire life cycle at the credit union.

FIGURE 11

IMPROVING DELINQUENCY PREDICTIONS



CHAPTER 4

How to Use Machine Learning

The data for this machine learning research come from five credit unions, all of which were interested in slightly different insights. Here we show how the same data can be used in very different ways to generate member insights—in this case, product recommendations, profitability predictions, and improved underwriting.

The Next Best Product

Using the cluster analysis described above, the next best product can be calculated with 30% accuracy. We tested the model on out-of-sample data points and found this prediction rate across all products and participating credit unions. **The next best product is the product that the member should be offered given his or her set of current products. The next best product is predicted by the member's switch to a different group and his or her transactional history.** For example, as was shown earlier in Figure 1, a member of cluster 1 is in general 15% likely to switch to cluster 9. Members in cluster 1 have a member share account, a nonredeemable member equity account, and a consumer term loan; 15% of these members will migrate to cluster 9, dropping the consumer term loan and getting a checking account.⁶ Thus, a checking account is the next best product for members of cluster 1, as they will very likely switch to this product in the near future.

In order to target the actual 15% of members who will move to cluster 9, however, the classifier described in Figure 4 must be calculated for each transaction and each member. When the classifier is above zero, the likelihood that the member will be interested in the new product set is approximately 30%.

Using this analysis, we were also able to predict the likelihood that members would leave the credit union and also which products they would start out with. For this analysis we looked at those clusters that were associated with almost no products, such as cluster 3 in Figure 3. At this credit union, in general members from cluster 3 switch to cluster 4, as shown in Figure 9. This indicates that members who have just joined the credit union very likely have a member share account (demand product) and a checking account (demand product).

In the case of one credit union, members in cluster 10 tend to leave the credit union. Interestingly, members in cluster 10 only have a member share account and are therefore only very loosely associated with the credit union before they leave.

Improving the Profitability and Wealth of Members

In our analysis we also incorporated the profitability and wealth of each cluster. In order to make the best possible suggestions to the member, we wanted to account for these factors in our recommendations.

Profitability was incorporated by calculating the difference in profitability between the clusters. Profitability was calculated using the fees and interest on credit or loans over savings in the product mix. Figure 12 shows the profitability as return on assets for the 10 example clusters shown in Figure 3.

As Figure 12 shows, members in clusters 2 and 4 are only slightly profitable. Members in cluster 1, however, return almost 1.50% on their assets for the credit union.

This analysis is extended even further by incorporating the wealth of each cluster. Wealth is incorporated by calculating the difference in average wealth between the clusters. The wealth of each cluster is calculated by the average savings minus average loans. Figure 13 shows the average wealth of each cluster shown in Figure 3. Clusters 6 and 10 have on average the lowest wealth in their member base. Members in cluster 1, however, have an average wealth of more than \$80,000.

Combining the profitability and wealth data of each cluster, we find that cluster 1 members are extremely profitable for the credit union and that its members are very wealthy.

We recommend that out of all the possible changes to the member's portfolio, the member be steered toward the option with the highest profitability and wealth; this can be calculated by adding the percentage difference in wealth and the percentage difference in profit.

FIGURE 12

AVERAGE PROFIT OF CLUSTERS 1-10 AT ONE CREDIT UNION

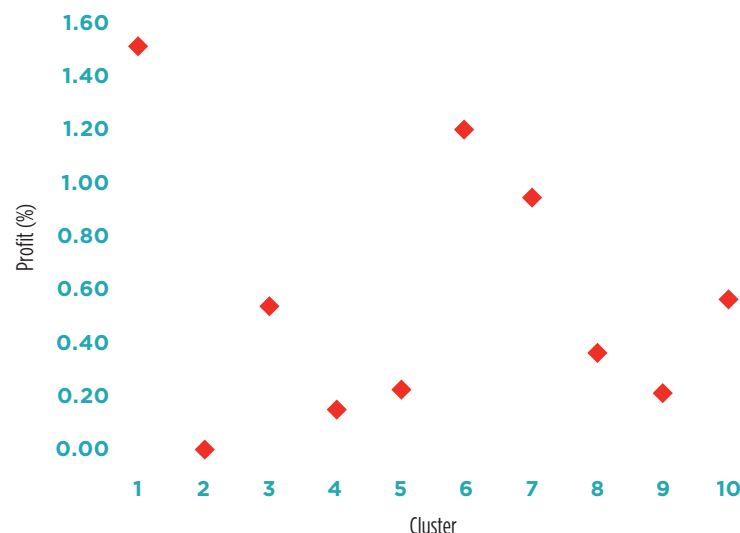
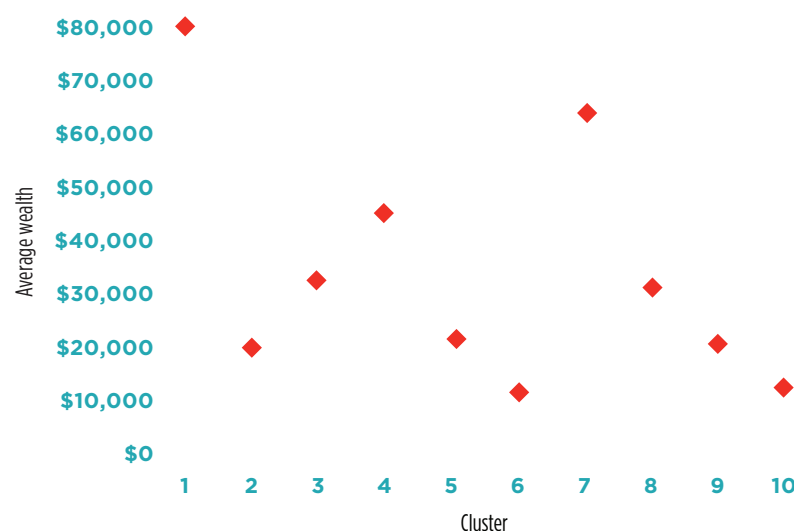


FIGURE 13

AVERAGE WEALTH OF CLUSTERS 1-10



Improving the Credit Scoring Model

The proposed machine learning model improved on the traditional credit scoring by about 10%, as shown in Figure 11. The model is an extension of the original credit scoring model in that it incorporates the original credit score as a factor but adds transactional variables that contribute to the prediction power. Based on this hybrid scoring model, we estimate that the participating credit unions can approve 1% more loans while reducing the cost of delinquencies by about 15%. In the provided data sets, the biggest gain could be achieved

among members with a credit score of between 450 and 600. This group has some of the lowest delinquency rates according to the original credit score, and that effect was even further enhanced by the hybrid scoring model. Therefore, this group might be a good target for credit cards, other consumer loans, and even mortgage products.

CHAPTER 5

Conclusion

In this report we analyzed patterns in transactional data from five participating credit unions that provided about 250 million transactions. The findings show that patterns evolve using machine learning and big data. In particular, we found that members follow some simple paths during their life cycle and adopt different consumer products at each stage.

For example, members move from simple products such as demand and savings accounts to revolving lines of credit and mortgages. These stages appear to be very predictable when looking at the transactions of each member. Before members add new products to their portfolio, in general their balance-to-income ratio increases or their income drops suddenly. These changes predict very accurately which kinds of consumer products might interest these members in the near future.

Before members add new products to their portfolio, in general their balance-to-income ratio increases or their income drops suddenly. These changes predict very accurately which kinds of consumer products might interest these members in the near future.

Each stage the member passes through predicts their adoption of several new consumer products. In our analysis we found that the best solutions are more profitable for the credit union and at the same time translate to higher wealth for the member. For example, lowering short-term interest rates might lead to more loans and therefore higher fees in the future. In some cases members switched from a revolving line of credit to a private loan. Of course, this reduces interest in the short run, but it can help the member to remain financially healthy and become more profitable in the future.

Another important finding is that external factors are a big influence on member behavior. For example, a financial crisis, indicated by the price of the S&P 500, has a high impact on income and is in general an early indicator of product changes. Some members impacted by the financial crisis through job loss and resulting lower income follow a pattern toward higher debt. This can be predicted by the credit union and supported with better structuring of the member's product portfolio.

Another factor that can be predicted through transactional data and external factors is credit risk. In this report we calculated the delinquency rate using a hybrid model and compared the predictions to those of the original credit scoring model currently applied by the participating credit unions. We compared the prediction accuracy and discrimination performance of the two models. In both cases, the hybrid was superior to the original credit scoring model. This suggests that credit risk, and ultimately the distribution of loans, should be based on additional variables such as the expense-to-income ratio or even external factors such as the recent price of the S&P 500.

This analysis is based on very extensive data sets and involves a large number of calculations. The most important finding is that members in all of the participating credit unions follow some very predictable patterns in their adoption of financial products. If these patterns are found, additional products can be promoted to members with a much higher probability of success. If these patterns are extended with information about the profitability and wealth of various groups, products can be suggested that improve the profitability and wealth of the member.

A prerequisite for developing these models is a well-maintained database with as many details about each transaction as possible. For example, some of the participating credit unions included information about the type of transaction and where it originated. These data are incredibly valuable, since member location and origin of transaction correlate highly with measures such as credit score and future financial products.

In order to conduct these analyses, several derived measures must be calculated using the transactional data, including expense-to-income ratio and total monthly inflow. These measures should be calculated through the transactional database for ease of use. Beyond that, the described model must be calibrated to the specific transactional variables by calculating the parameters of the prediction algorithm—in particular the weights of the input variables.

One possible extension of the described method would be to include additional external variables beyond the S&P 500. Additional variables might include information about other behaviors that predict future financial performance. The use of a car predicts higher costs due to car insurance and maintenance, which ultimately lowers usable income.

Health-related behavior is another predictor of credit default, as shown by the Jumiya project. In particular, an active lifestyle predicts a lower probability of defaulting or being late on payments. An active lifestyle lowers the chances of obesity and heart disease, and medical debt accounts for more than 60% of all bankruptcies in the United States. Given the close relationship between health-related behavior and financial health, information on health-related behavior could improve predictions of a member's credit risk and future financial products.

FIGURE 14

ANALYZED CREDIT UNION PRODUCTS

Product	Type	Description
shares	Demand	Member share account
cheq	Demand	Checking account
conv	Demand	Convenience Plus account
pat	Demand	Nonredeemable member equity account
T5	Term	5-year term
rpat	Demand	Redeemable member equity account
max	Demand	Maximizer account
Consum	Loan	Consumer term loan
RSPVAR	Demand	Variable RSP
Qline	Loan	QuickLine
DDA	Demand	Deposits
SAV	Demand	Savings
RC	Loan	Revolving line of credit
CD	Term	Term deposits
MTG	Loan	Personal mortgages
LN	Loan	Loans
COM	Loan	Commercial mortgages
SV01	Demand	Regular
B01	Demand	Basic

Product	Type	Description
SV05	Demand	05 — Regular
SV02	Demand	02 — Regular
SK01	Demand	SeekMore Checking
SV03	Demand	03 — Regular
SV04	Demand	04 — Regular
IS01	Loan	IRA savings / IRA Accumulator
1051	Term	12 Month (Silver)
1031	Term	6 Month (Bronze)
A-RSP	Demand	RSP demand savings
B	Demand	Demand
C	Demand	Demand
H-TFA	Demand	TFSA demand savings
J	Demand	Demand checking USD
M	Demand	Demand
N	Demand	Demand savings
FRM-53	Loan	Mortgage, variable rate
MTG-23	Loan	Mortgage, fixed rate
K-RIF	Term	RIF term

Endnotes

¹ See, for example, Amir E. Khandani, Adlar J. Kim, and Andrew W. Lo, “Consumer Credit-Risk Models via Machine-Learning Algorithms,” *Journal of Banking & Finance* 34, no. 11 (2010): 2767–87.

² Harland Financial Solutions offers software that performs a similar function.

³ See Khandani, Kim, and Lo, “Consumer Credit-Risk Models,” for more information on these methods.

⁴ See the appendix for product descriptions and types.

⁵ David U. Himmelstein, Deborah Thorne, Elizabeth Warren, and Steffie Woolhandler, “Medical Bankruptcy in the United States, 2007: Results of a National Study,” *American Journal of Medicine* 122, no. 8 (2009): 741–46.

⁶ See the appendix for product descriptions.

List of Figures

9	FIGURE 1	NEXT BEST PRODUCT ANALYSIS
10	FIGURE 2	EXAMPLE OF <i>k</i> -MEANS CLUSTERING
11	FIGURE 3	RESULTS OF THE CLUSTER ANALYSIS
11	FIGURE 4	DISTANCE MEASURE
12	FIGURE 5	INPUT VARIABLES
13	FIGURE 6	S&P 500 PRICE, 2003–2013
13	FIGURE 7	STORE CHOICE AS PREDICTOR OF MEMBER CREDIT SCORE
14	FIGURE 8	MOST LIKELY TO SWITCH
14	FIGURE 9	NEXT BEST CLUSTER ANALYSIS
15	FIGURE 10	INPUT FILTER
16	FIGURE 11	IMPROVING DELINQUENCY PREDICTIONS
18	FIGURE 12	AVERAGE PROFIT OF CLUSTERS 1–10 AT ONE CREDIT UNION

18	FIGURE 13 AVERAGE WEALTH OF CLUSTERS 1–10
22	FIGURE 14 ANALYZED CREDIT UNION PRODUCTS

About the Author



Philipp Kallerhoff, PhD

Jumiya

Philipp Kallerhoff studied engineering, management, psychology, and physics at the Technical University Berlin, Sophia University in Tokyo, and the State University of New York. After completing a PhD in computational neuroscience at the Technical University Berlin, he worked in the banking and hedge fund industries. He completed Singularity University's Graduate Studies Program, sponsored by Google and NASA, and in 2012 cofounded Jumiya, a free rewards program that helps people get healthier and wealthier.

About Filene

Filene Research Institute is an independent, consumer finance think and do tank. We are dedicated to scientific and thoughtful analysis about issues affecting the future of credit unions, retail banking, and cooperative finance.

Deeply embedded in the credit union tradition is an ongoing search for better ways to understand and serve credit union members. Open inquiry, the free flow of ideas, and debate are essential parts of the true democratic process. Since 1989, through Filene, leading scholars and thinkers have analyzed managerial problems, public policy questions, and consumer needs for the benefit of the credit union system. We support research, innovation, and impact that enhance the well-being of consumers and assist credit unions and other financial cooperatives in adapting to rapidly changing economic, legal, and social environments.

We're governed by an administrative board made up of credit union CEOs, the CEOs of CUNA & Affiliates and CUNA Mutual Group, and the Chairman of the American Association of Credit Union Leagues (AACUL). Our research priorities are determined by a national Research Council comprised of credit union CEOs and the President/CEO of the Credit Union Executives Society.

We live by the famous words of our namesake, credit union and retail pioneer Edward A. Filene: "Progress is the constant replacing of the best there is with something still better." Together, Filene and our thousands of supporters seek progress for credit unions by challenging the status quo, thinking differently, looking outside, asking and answering tough questions, and collaborating with like-minded organizations.

Filene is a 501(c)(3) not-for-profit organization. Nearly 1,000 members make our research, innovation, and impact programs possible. Learn more at filene.org.

"Progress is the constant replacing of the best there is with something still better!"

—Edward A. Filene



612 W. Main Street
Suite 105
Madison, WI 53703

p (608) 661 3740
f (608) 661 3933

Publication #301 (6/13)