

# Predicting Customer Behavioural Patterns using a Virtual Credit Card Transactions Dataset

Ayrton Senna Azzopardi and Joel Azzopardi

*Department of Artificial Intelligence, Faculty of ICT, University of Malta, Msida, MSD2080, Malta*

**Keywords:** Data Mining, Financial Transactions, Customer Churn Prediction.

**Abstract:** Nowadays, many businesses are resorting to data mining techniques on their data, to save costs and time, as well as to understand customers' needs. Analysing such data can lead to higher profits and higher customer satisfaction. This paper presents a data mining study that is applied on millions of transactional records collected for a number of years, by a leading virtual credit card company based in Malta. In this study, 2 machine learning techniques, namely Artificial Neural Networks (ANN) and Gradient Boosting (GBM), are analysed to identify the best modelling framework that predicts the churning behaviour of this company's customers. Apart from helping the marketing department of this firm by providing a model that predicts churning customers, we contribute to literature by identifying the minimum amount of customer activity needed to predict churn. In addition, we also analyse the "cold start" problem by performing a time-series experiment based on the few data available at the beginning of the customer purchase history.

## 1 INTRODUCTION

With the advancements in web technologies, online shopping as well as online gambling have rapidly increased in popularity in the past years. In fact, worldwide retail e-commerce sales are showing logarithmic growth and forecasted to exceed 5.4 trillion USD (\$) in 2022 and 6.3 trillion USD in 2024<sup>1</sup>, whilst the online gambling market is forecasted to exceed 92.9 billion USD in 2023<sup>2</sup>. The ability of making a financial transaction instantly and from across the world is a core reason why e-commerce and gambling websites grew at such a rapid pace. However, when making online transactions expose users' financial data. This has been a major concern to some users, and so they are increasingly resorting to methods that help protect their financial information.

One such popular approach is through virtual credit cards. Funds are initially deposited into an account linked with such a virtual credit card, and then the actual virtual credit card is used when performing transactions online. In this way the actual bank account details of a user are not exposed when making a transaction on the web. Moreover, the actual bank

will not have any knowledge whatsoever how clients are spending their money. This is extremely beneficial to online gamers as certain banks are rejecting attempts of credit cards used in gambling websites.

The increase in popularity of e-commerce and gambling websites is leading to millions of users using virtual credit card services. Consequently, virtual credit card companies are dealing with huge amounts of incoming data in the form of new users, creation of credit cards, deposit of funds and mostly online transactions. The customer data has the potential to be mined in order to extract meaningful knowledge about the different users and any trends that might lead to increased profits and customer satisfaction.

For this reason, we present a data mining study that is applied on millions of transactional records that were collected for a number of years, by a leading virtual credit card company based in Malta. We utilise machine learning techniques to predict the churning behaviour of this company's customers. We also analyse and identify the minimum amount of customer activity needed to predict churn. In addition, we attempt to handle the "cold start" problem by performing a time-series experiment based on the minimal data available at the beginning of a customer's purchase history.

<sup>1</sup>May 2022: <https://www.trade.gov/e-commerce-sales-size-forecast>

<sup>2</sup>May 2022: <https://www.statista.com/statistics/270728/market-volume-of-online-gaming-worldwide/>

## 1.1 Aims and Objectives

The aim of this research is to investigate machine learning techniques and develop a modelling framework that is capable of predicting whether a customer is churning based on one's virtual credit card transactions. The aforementioned aim will be fulfilled through the following objectives:

- Extract a number of dynamic features from the provided raw financial transactions, so as to find the most effective feature set when predicting customer churn.
- Build a machine learning setup that is effective in modelling and predicting customer churn using the extracted features.
- Determine the minimum amount of customer activity needed, in order to effectively predict whether a customer is churning or not.
- Determine whether demographic features together with any initial financial transactions can be used to predict whether a relatively new customer will continue to use the company's services or else churns.

## 1.2 Contribution to Research

Despite the fact that there are quite a few researchers to have studied customer churn in the financial domain, the concept of predicting churn using dynamic features solely extracted from raw financial transactions, as presented in this paper, is still quite a novel approach. Traditional customer churn prediction solutions focus solely on domain specific and static features. Following recent research that have shown the effectiveness of dynamic behavioural predictors, we extract meaningful dynamic features from a customer's virtual credit card usage history, and validate their usefulness with regards to customer churn prediction.

Our second objective is to construct a machine learning setup that is able to model the churn behaviour of customers. Section 2 provides review of a number of different machine learning techniques applied for this problem. Unfortunately, one can not that the generated framework with the best results is specific to the dataset that is used in that particular experiment. It does not entail that the same maintain the same performance when applied on different datasets. For this reason, our second objective evaluates and finds the best performing technique on the dataset that we utilise that records virtual credit card transactions.

Our third objective attempts to identify the optimal observation window size. From all the reviewed

literature, only Leung and Chung (Leung and Chung, 2020) attempt to test different observation window sizes. However, this experiment was solely an evaluation of just 2 observation periods differing in length (4 months vs 6 months). We fill this research gap by performing a more extensive time-series experiment so as to determine the minimum amount of customer spending history needed.

Finally, in our fourth objective we tackle the "cold start" problem. This objective is mainly motivated by the fact that since organisations would barely have any observed data for newly registered customers, the latter are generally excluded from further analysis. In this research, we attempt to fill this gap by applying another time-series experiment using customer demographics – the only data available at that specific time.

## 2 LITERATURE REVIEW

The financial and banking industry has been evolving quite substantially in recent years. Consequently, existing companies in this industry are facing extreme competition, not only by direct competitors but also by new entrants and start-ups that can be disruptive by providing innovative financial solutions (Shirazi and Mohammadi, 2019). Different industries are focusing more on managing the churn behaviour of customers, rather than investing in strategies in an attempt to acquire new customers given the fact that customer retention is far less costly than acquiring new customers (Kaya et al., 2018; Bilal Zorić, 2016; Kim et al., 2005; Rosa, 2019; Shirazi and Mohammadi, 2019; Safinejad et al., 2018; Leung and Chung, 2020; Szmydt, 2018; Keramati et al., 2016; Farquard et al., 2014).

In this section, we will go through some of the work found in literature with regards to predicting the churn behaviour of customers in the financial sector, giving an overview of the decisions taken by academics during their research when tackling the extraction of churn related features, the modelling of their machine learning approach and finally the choice of the observation window size.

Kim et al. (Kim et al., 2005) are amongst the first academics that started researching this problem. They applied Support Vector Machines (SVM) to the analysis of customer churn behaviour and evaluate its effectiveness through the use of demographic and credit card usage information. They only utilise a 3-month period when observing the credit card usage of customers as they claim that such period is adequate in understanding the behaviour of a customer. Their model is compared against a three-layer Back-Propagation Neural Network (BPN), where SVM ob-

tained better prediction results. In addition, Kim et al. (Kim et al., 2005) state that the process of parameter tuning is a vital step since different parameter values drastically change the prediction performance.

Similarly, Farquad et al. (Farquad et al., 2014) also apply SVM to predict customer churn. However, they go a step further by constructing a hybrid model, where apart from predicting whether a customer is churning or not, the model is capable of extracting informative rules on the customers. Their approach can be viewed in 3 phases. Initially, the number of features used when predicting churn is reduced through a recursive feature elimination process. Hereafter, the support vectors computed after training the SVM with the reduced feature set, are extracted. These support vectors together with the predicted values are used to construct a new dataset to be utilised in the final phase. Eventually, a Naive-Bayes Tree is implemented to purposely generate meaningful rules giving more insights about the churn behaviour of customers.

One can find a number of works in literature that follow a similar approach to the one applied in (Farquad et al., 2014), where researchers aim to generate a number of informative rules when analysing customer churn. The generated rules tend to group customers into different segments according to common behaviour. For instance, Keramati et al. (Keramati et al., 2016) aim to outline common characteristics of churning customers. To identify any hidden behavioural patterns, a Decision Tree (DT) is employed since, by nature, DT generate clear and significant “if-then” rules, allowing the authors in (Keramati et al., 2016) to fulfil their aim. Similarly, Cil et al. (Cil et al., 2018) also utilise DT in their quest to discover meaningful knowledge from their dataset. Utilising a dataset consisting of socio-demographic information together with nearly 4,000,000 investment fund transactions of around 65,525 bank customers they analyse up to 6 months prior to the closure or inactivity of a customer’s account from the customer’s entire investment fund transaction history. Subsequently, they perform 2 types of analysis. Initially, they model a DT on the computed investment fund transaction order data so as to determine the transactional patterns of customers that closed off their account with the bank. The learnt patterns in the form of DT rules are then utilised on future customers to predict those that potentially are willing to churn.

On the other hand, other studies do not focus on acquiring such rules to characterise churning customers, but rather focus on predicting whether a customer is churning or not as efficiently as possible. A popular approach in this case is the standard ANN. Bilal (Bilal Zorić, 2016) proposes a neural network

based framework that is capable of predicting the likelihood of churn for customers of a small Croatian bank. The dataset used in this study contains merely socio-demographic information and levels of service usage. Similarly, (Safinejad et al., 2018) employ a non-linear ANN to predict future churn rate of financial customers. They make use of a dataset that contains raw financial transactions of more than 4,500 customers recorded between 2009 and 2011. The 3-year observation window is divided into seasonal intervals and for each interval (12 in total), Recency (R), Frequency (F), Monetary (M) and Length (L) variables are calculated as features. Subsequently, they utilise a “fuzzy dynamic model” that can be split into 3 phases. Firstly, a weighted-RFML model is utilised to cluster customers and identify the segment representing the most valuable customers. Secondly, a fuzzy rule-based model that takes as inputs the L, F and M variables and outputs a 3-mode (low, medium, high) churn rate value, is developed. Thirdly, the prediction of future churn rate is modelled using 2 models - ARIMA as a linear machine learning model and ANN as a non-linear model. The authors in (Safinejad et al., 2018) conclude that the ANN model outperformed the other while claiming to have identified a suitable model for customer churn prediction together with an appropriate definition of churn in the finance sector.

Most of the customer churn prediction systems reviewed above, focus solely on domain specific and static features. Some of these features that are extracted and used in such traditional attempts, generally represent product or account type ownership, service usage aggregation and socio-demographic information. Dynamic behavioural patterns in a customer’s financial transactions, are rarely considered.

In fact, Kaya et al. (Kaya et al., 2018) try to fill this research niche by exploring the spatio-temporal patterns and choice behaviour of customers and determine whether such behaviour relates to the customer churn event. They extract novel features based on the spatio-temporal and choice patterns of customers. Spatio-temporal features include *diversity*, *loyalty* and *regularity* that measure how varied or constant customers are within their purchase behaviour with regards to time and location perspective. On the other hand, the financial choice patterns outline how customers disperse their spending with regards to merchants, purchase categories and locations of merchants. (Kaya et al., 2018) claim that churn activities can be effectively predicted using dynamic behavioural patterns and furthermore, using domain-independent variables specifically those that are based on the spatio-temporal patterns in human activities.

Leung and Chung (Leung and Chung, 2020) also propose a dynamic classification framework that utilises actual customer behavioural patterns. For every customer, the utilised dataset contains static predictors, such as the traditional demographic and product ownership variables, and also account activity predictors including aggregation of financial transactions and service usage. A trend factor is computed for each account activity predictor, aiming to capture the trends of account activities within the observation period. The authors in (Leung and Chung, 2020) experiment with both the observation window (4 months vs 6 months) and the labelling window (2 months vs 3 months). After computing the trend factors for each different period, they evaluate 3 different supervised machine learning models, namely Logistic Regression (LR), Random Forest (RF) and Gradient Boosting (GBM). The authors conclude that with 6 months of data, the models obtained better accuracy than with 4 months of data. On the other hand, accuracy decreases rapidly as the prediction window is extended. Furthermore, RF and GBM outperformed the LR prediction model.

Deep learning models are known to be able to capture high-level representations from the huge amounts of customer data being created at an increasing rate due to the increasing amount of financial activities (Hsu et al., 2019). In fact, Hsu et al. (Hsu et al., 2019) develop a Recurrent Neural Network (RNN) feature extractor with GRU. The aim is to better model the time dependencies found within a customer's credit card spending history, as a result of which, a number of dynamic features are then extracted for customer churn prediction. They propose an innovative approach by combining this strong dynamic feature extraction from RNN with a RF. This enhanced RNN-RF model is therefore capable of combining dynamic and static features allowing for better performance when predicting credit card customer churn. They evaluate their innovative approach on a dataset consisting of 30,000 instances with 23 features each, separated into 5 static socio-demographic features and 18 dynamic features describing monthly the customer's service usage in a 6 month period. The authors conclude that the RNN-RF predictive model outperformed other benchmark models and also stated that the model performed better with more training instances. Further analysis on the use of DL models for this problem is provided in (Jain and Jayabalan, 2022).

### 3 METHODOLOGY

The objectives discussed in Section 1.1, were achieved by constructing a customer churn predictive framework that can be applied on raw financial transactions.

Initially, this framework pre-processes the raw financial transactions to filter out any missing or redundant data. In the pre-processing stage, the financial transactions are also segmented into several time periods as a preparatory work for future analysis. Hereafter, numerous features are generated from the financial transaction records so as to extract and represent any behavioural patterns hidden within the data. Furthermore, customers are also classified and labelled according to our customer churn definition. Subsequently, the extracted features and the generated customer churn labels are fed into a machine learning technique in order to model the churn behaviour of customers and be able to predict churn activities.

Consequently, after determining the best set of features and the best performing predictive model, an experiment is performed where different observation window sizes are evaluated to determine the minimum amount of customer activity required prior to the churn event. Finally, a time-series experiment is performed to determine whether demographic information combined with any spending information that is available at the beginning of the customer's relationship with the company, can be used to predict whether new customers are willing to continue using the company's service.

#### 3.1 The Extracted Features

In this study, we extracted different types of features aimed at representing the customer behaviour required for churn prediction.

In our review of literature, we have seen how the authors in (Cil et al., 2018; Bilal Zorić, 2016; Kim et al., 2005; Rosa, 2019; Shirazi and Mohammadi, 2019; Leung and Chung, 2020; Hsu et al., 2019; Leung and Chung, 2020; Keramati et al., 2016; Farquard et al., 2014) all make use of socio-demographic information within their churn predictive systems. For this reason, we followed their approach and made use of the customer demographics found in the provided dataset as well. Furthermore, we have seen how the authors in (Kim et al., 2005; Keramati et al., 2016; Bilal Zorić, 2016; Rosa, 2019; Leung and Chung, 2020) also make use of static predictors representing the service usage of customers. With this in mind, we compute a number of statistical features that aggregate different aspects of the purchase history of a cus-

customer within a particular observation window. These are referred to as “global” statistics.

Latest research has shown that dynamic behavioural features tend to be more effective in representing customer behaviour for churn prediction (Kaya et al., 2018; Leung and Chung, 2020; Hsu et al., 2019). For this reason, some of the “global” statistics computed on the entire observation window, were also applied for each month period in the window, creating a new set of macro-average monthly statistics. Such feature extraction process is similar to the one performed by Leung and Chung (Leung and Chung, 2020). However they then compute a trend factor value based on the computed dynamic features, prior to inputting the features into the predictive model. At this stage, we followed the approach taken by Kaya et al. in (Kaya et al., 2018) and inputted the dynamic features directly in the predictive model. In addition, we also computed a feature vector containing the amounts spent by the customer on each day of the observation window. The motivation behind this, is to allow the predictive model to train on variables that resemble the raw financial transactions as much as possible in an attempt to not lose any dynamic behavioural information of customers.

Finally, inspired by the innovative idea of having features representing the choice behaviour of customer with regards to merchants, purchase categories and locations of merchants (Kaya et al., 2018), we generated a merchant vector containing the number of purchases done towards each Merchant Category Code (MCC).

As can be seen, the constructed framework utilises different types of features including demographic, static and dynamic predictors so as to predict the churning customers in the financial industry.

### 3.2 Machine Learning Techniques

The focus of our framework is not to acquire rules and determine common characteristics of churning customers, but rather to efficiently predict whether a customer is churning or not. For this reason, we followed the approach taken by (Bilal Zorić, 2016; Rosa, 2019; Safinejad et al., 2018), and employed a Neural Network (ANN) as our predictive model. ANNs are intended to artificially replicate the behaviour of the biological systems found in the human brain. For this reason, similar to other researchers contributing to the field of churn prediction, we believed that this non-linear predictive model could be a suitable contender in modelling the churn behaviour.

Furthermore, we also employed a Gradient Boosting Model (GBM) as our customer churn predictive

model. The fact that GBM is capable of tuning weak predictive models so as to become better predictors by generating a single predictive model as an ensemble of numerous weak ones, inspired us to make use of such technique in our quest to predict customer churn. In addition, (Leung and Chung, 2020) stating that GBM outperformed a LR and was also on par with a RF, gave us further motivation in constructing such predictive model.

Both classifiers were implemented using readily available libraries. We decided to implement our ANN using Keras<sup>3</sup> which is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. On the other hand, we decided to construct our GBM using XGBoost<sup>4</sup>. XGBoost is an optimised distributed gradient boosting library designed to be highly efficient, flexible and portable.

### 3.3 Data Observation and Labelling

Most of the studies reviewed in Section 2, did not opt for an observation window exceeding 6 months. In fact, most of the systems we reviewed employ an observation window size varying between 3 to 6 months. As a result, we performed an experiment where we train our predictive model on varying observation window sizes starting from 1 month worth of data up to 6 months. The results of such experiment will give us an inclination of the minimum amount of observed customer activity needed to predict churn.

On the other hand, when we observed the financial transactions in the period following the observation window so as to determine the churn label for customers, we decided to only consider the succeeding month. This is mainly because according to Leung and Chung (Leung and Chung, 2020), prediction accuracy decreases instantly as the prediction window increases. Furthermore, companies and marketing departments would find it more beneficial if they can predict what activity is expected in the coming month. In addition, we did not employ any fuzzy logic in our customer churn definition, meaning that a customer can either be labelled as “Churned” or “Not Churned”. In fact, our customer churn definition is quite straightforward - if a customer has at least 1 transaction in the labelling window then the label is “Not Churned” else “Churned”.

<sup>3</sup>May 2022: <https://github.com/keras-team/keras>

<sup>4</sup>May 2022: <https://github.com/dmlc/xgboost/tree/master/python-package>

### 3.4 The Cold Start Problem

In an attempt to address the “cold start” problem in machine learning problems, a time-series experiment is performed where we initially try to predict the churning behaviour of newly registered customers using only their demographic information. Then, as new data starts coming in for these customers, we added their daily purchase amounts to the feature set and performed the customer churn prediction once again, measuring the performance each time. This time-series experiment was performed on a weekly basis, meaning that the churn prediction was performed whenever another week of the customer’s purchase history was observed, each time adding a feature vector of length 7, comprising of the purchase amounts for every day in that week.

## 4 EVALUATION

The four research objectives set in Section 1.1, so as to fulfil the project’s aim, are measured and evaluated, as discussed in the following subsections.

### 4.1 Evaluation of Similar Systems

One of the main challenges that is encountered by research within the financial domain, is the sensitive nature of such financial data. According to (Martens et al., 2016), studies on data processing and analysis for financial businesses, highly depend on close collaborations with the industry (Cil et al., 2018; Safinejad et al., 2018). However, any company data together with the information resulting from the research rarely get shared with the scientific community, due to its sensitive nature. In light of such issue, evaluation is challenging due to the lack of gold standard datasets available. In fact, all the related systems reviewed in Section 2, do not compare their findings against those obtained in other research work, but solely evaluate their own proposed solution on a dedicated test set and discuss the results in terms of various performance metrics.

Apart from the traditional Accuracy score, the performance of most classification models is measured using the AUROC metric. This metric provides an aggregate measure of performance across all possible classification thresholds. This metric has been used to evaluate the systems described in (Keramati et al., 2016; Rosa, 2019; Kaya et al., 2018; Hsu et al., 2019).

### 4.2 Evaluation of the Extracted Features and the Machine Learning Techniques Used

In order to determine the most effective features in capturing customer behavioural patterns and the best performing modelling technique in customer churn prediction, a greedy search was employed. We computed all the combinations of the different feature categories and fed the computed feature sets into our two predictive models, measuring the prediction performance of each model using the AUROC metric. It is worth mentioning, that the observation window for this evaluation experiment was taken to be 3 months.

The ANN classification model managed to obtain its highest AUROC score when trained on demographic information, global statistics aggregating the entire 3-month observation window and the vector comprising of the daily purchase amounts, obtaining a score of 0.62. With such performance, we can conclude that the implemented ANN model had some form of ability in distinguishing between the 2 classes. On the other hand, the GBM classification model performed at its best when trained on global statistics acquired from the entire observation window, monthly statistics obtained for each month in the window, the vector consisting of the number of purchases done towards the different MCCs and finally the vector comprising of the daily purchase amounts. In this scenario, the GBM managed to obtain an AUROC score of 0.69, distinguishing the 2 classes way better than the implemented ANN model. Furthermore, we noticed that the GBM improved its score by a few percentages whenever more features are observed.

In addition, for the best performing ANN and the best performing GBM setup, we computed other performance metrics using traditional values from the confusion matrix so as to have a better understanding of the models’ prediction performances. In fact, we computed the Sensitivity, Specificity, False Positive Rate, False Negative Rate and the Precision metric. The obtained measurements are shown in Tables 1 and 2 for the ANN and GBM respectively.

Table 1: Evaluation results of the ANN predictive model based on the Confusion Matrix.

Metric	Score
Sensitivity	0.4535
Specificity	0.7284
False Positive Rate	0.2716
False Negative Rate	0.5465
Precision	0.6238

Table 2: Evaluation results of the GBM predictive model based on the Confusion Matrix.

Metric	Score
Sensitivity	0.6989
Specificity	0.6865
False Positive Rate	0.3135
False Negative Rate	0.3011
Precision	0.6581

From these metrics, one can conclude that the GBM with 70% sensitivity, was more capable of predicting positive cases i.e. “Churners”, whilst on the other hand ANN with 73% specificity, was more capable of identifying negative cases i.e. “Non Churners”. Furthermore, GBM was quite consistent and managed to incorrectly classify both positive and negative cases around 30% of the time. On the other hand, ANN incorrectly classified positive cases as negative around 55% of the time. With regards to how many predicted positive cases were actually correct, the GBM edged the ANN model with a 3.5% better precision. To conclude this section, it is fair to say that the implemented GBM is more suitable to predict the churn behaviour of customers. In view of this, the constructed GBM model and all the extracted features bar customer demographics, are utilised in the remaining experiments.

### 4.3 Evaluation of the Varying Observation Window Sizes

In this section, we examined different observation window sizes and checked how the prediction performance of the classification model changes in return, so as to determine the minimum amount of customer purchase history required while still predicting churn with the same performance. The results of such experiment are shown in Table 3.

Table 3: Evaluation results of the different observation window sizes.

Number of Months	AUROC Score
1	0.6674
2	0.6883
3	<b>0.6927</b>
4	0.6865
5	0.6801
6	0.6850

Despite varying the observation window size from 1 month up to 6 months, the performance of the predictive model did not change extensively however. The 3-month observation window remained with the best performance metric score, with the other window

sizes not managing to cap that. Despite obtaining the lowest AUROC score (0.67) the 1-month observation window was only around 2.5% off the top. It can be concluded that decreasing the amount of observed purchase history is possible without sacrificing too much predictive performance.

### 4.4 Evaluation of the Usage of Demographics and any Initial Observations for Churn Prediction on New Users

In this section, we conducted a time-series experiment where we examined how effective the classification model is when predicting churn on new customers. The results of such experiment are shown in Table 4.

Table 4: Evaluation results of the usage of demographics combined with the initial purchase observations for churn prediction on new users.

Number of Weeks	AUROC Score
0	0.5026
1	0.5746
2	0.6175
3	0.6524
4	<b>0.6828</b>

Results show that knowing just the age, country and currency information of a customer, is not enough to be able to predict whether a newly registered user is willing to continue using the company’s services or rather stops and defaults. Predicting customer churn with only demographic data is as effective as tossing a coin. It is worth noting that with a few weeks of observed purchase data, the prediction performance increased quite rapidly, reaching the levels of having a 3-month observation of purchases. Both this experiment and the one preceding it, have shown that a month’s worth of data is still quite sufficient to predict whether a customer is defaulting in the next month. We can conclude that by only observing the purchase data of the current month, we can infer churn predictions for the following month.

## 5 CONCLUSION AND FUTURE WORK

In this paper, we presented a data mining study that was applied on millions of financial transactions collected for a number of years, by a leading virtual credit card company based in Malta.

All 4 objectives specified in Section 1.1 have been

fulfilled. With regards to the first objective (identifying the most effective feature set), we extracted different features, namely: demographic features, “global” statistics that are relative to the entire observation window, dynamic monthly statistics for each month in the observation window, a vector containing daily purchase amounts and another vector containing the number of purchases done towards each MCC. The best results were achieved when utilising all the features except for demographic features.

We fulfilled the second objective (build a setup that can model and predict customer churn) by applying the Artificial Neural Network and the Gradient Boosting Model for this problem. The GBM classifier resulted in the best machine learning framework of this study, obtaining an AUROC score of 0.6927. In addition, we also observed that our learning framework is capable of correctly identifying 70% of “Churners”, potentially making it a suitable solution in Customer Relationship Management.

When handling the third objective (to determine the minimum amount of customer activity needed in order to predict its likelihood of churning), experiment results show that decreasing the observation window to a month’s length does not extensively affect the predictive performance of the classifier, giving the ability to negotiate between prediction accuracy and amount of data observed.

For our final objective (attempting to handle the cold-start problem using a customer’s demographic features that can be made accessible upon registration), we attempted to predict customer churn using only demographic information and in time, combine any new purchase data. This experiment showed that for the current dataset, predicting churn behaviour using only customer demographics (the customer’s age, country and currency information), is not anywhere sufficient enough to be able to predict whether a newly registered customer is going to default or not in the coming month.

The work described in this paper can be further improved by augmenting the constructed framework to a tree-based model in order to extract meaningful behavioural rules. These can be used to capture the actual characteristics of churning customers. Furthermore, after addressing the problem of customer churn prediction, it now makes sense to tackle the problem of predicting the next purchases of customers. The approaches performed in collaborative recommendation systems can be adopted and tweaked to our purpose.

## REFERENCES

- Bilal Zorić, A. (2016). Predicting customer churn in banking industry using neural networks. *Interdisciplinary Description of Complex Systems: INDECS*, 14(2):116–124.
- Cil, F., Cetinyokus, T., and Gokcen, H. (2018). Knowledge discovery on investment fund transaction histories and socio-demographic characteristics for customer churn. *International Journal of Intelligent Systems and Applications in Engineering*, 6(4):262–270.
- Farquard, M. A. H., Ravi, V., and Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical crm application. *Applied Soft Computing*, 19:31–40.
- Hsu, T.-C., Liou, S.-T., Wang, Y.-P., Huang, Y.-S., et al. (2019). Enhanced recurrent neural network for combining static and dynamic features for credit card default prediction. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1572–1576. IEEE.
- Jain, S. V. and Jayabalan, M. (2022). Applying machine learning methods for credit card payment default prediction with cost savings. *Biomedical and Business Applications Using Artificial Neural Networks and Machine Learning*, pages 285–305.
- Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B., et al. (2018). Behavioral attributes and financial churn prediction. *EPJ Data Science*, 7(1):41.
- Keramati, A., Ghaneei, H., and Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, 2(1):10.
- Kim, S., Shin, K.-s., and Park, K. (2005). An application of support vector machines for customer churn analysis: Credit card case. In *International Conference on Natural Computation*, pages 636–647. Springer.
- Leung, H. C. and Chung, W. (2020). A dynamic classification approach to churn prediction in banking industry.
- Martens, D., Provost, F., Clark, J., and de Fortuny, E. J. (2016). Mining massive fine-grained behavior data to improve predictive analytics. *MIS quarterly*, 40(4).
- Rosa, N. B. d. C. (2019). *Gauging and foreseeing customer churn in the banking industry: a neural network approach*. PhD thesis.
- Safinejad, F., Noughabi, E. A. Z., and Far, B. H. (2018). A fuzzy dynamic model for customer churn prediction in retail banking industry. In *Applications of Data Management and Analysis*, pages 85–101. Springer.
- Shirazi, F. and Mohammadi, M. (2019). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*, 48:238–253.
- Szmydt, M. (2018). Predicting customer churn in electronic banking. In *International Conference on Business Information Systems*, pages 687–696. Springer.