Gauging and Foreseeing Customer Churn in the Banking Industry:
A Neural Network Approach

Nelson Belém da Costa Rosa

Project Work report presented as partial requirement for obtaining the Master’s degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa
GAUGING AND FORESEEING CUSTOMER CHURN IN THE BANKING INDUSTRY:

A NEURAL NETWORK APPROACH

by

Nelson Rosa

Project Work report presented as partial requirement for obtaining the Master’s degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor: Prof. Roberto Henriques, PHD

November 2018
ABSTRACT

In today’s highly competitive market where information is easily accessible and with ever decreasing switching costs, customers in any industry hardly hesitate to move their business elsewhere should they find more suitable proposals to accommodate their needs. As such, it is in the best interest of any company to keep a close watch on its customers to monitor any signs of potential churning down the line. Significant headways in the Business Intelligence field have brought forward a great number of tools for knowledge discovery and predictive analytics purposes.

This paper proposes a new framework for assessing and predicting customer attrition in one of the biggest Portuguese retail banks. Data mining techniques were employed to study the behavior and patterns of past churners. A set of predictor variables was obtained from this process and used to train a set of predictive models using neural networks. Assessment of the performance of the classifiers was later validated on a sample of current customers in risk of churning.

The goal of this project is to provide a new approach to identify potential churners so marketing retention strategies be developed accordingly.

KEYWORDS

Banking; Business Intelligence; Customer Churn; Data Mining; Neural Networks
INDEX

1. INTRODUCTION .................................................................................................................. 1
   1.1 – Background and Problem Identification ...................................................................... 1
   1.2 – Study Objectives .......................................................................................................... 2
   1.3 – Study Importance .......................................................................................................... 3

2. LITERATURE REVIEW ....................................................................................................... 4
   2.1 – Definition of Churn ...................................................................................................... 4
   2.2 – Overview of the Retail Banking Industry and Effects on Churn ............................. 5
   2.3 – A Study of Past Churners ............................................................................................ 7
   2.4 – Churn Prediction in Banking ......................................................................................... 11
   2.5 – Artificial Neural Networks ......................................................................................... 17
   2.5.1 – Single Layer Perceptron .......................................................................................... 17
   2.5.2 – Multi-Layer Perceptron .......................................................................................... 18
   2.5.3 – The Backpropagation Algorithm ............................................................................ 19

3. METHODOLOGY ................................................................................................................ 21
   3.1 – Business Understanding .............................................................................................. 22
   3.2 – Data Understanding ..................................................................................................... 22
   3.3 - Data Preparation ......................................................................................................... 25
   3.4 - Data Pre-Processing .................................................................................................... 27
      3.4.1 - Data Cleansing ....................................................................................................... 27
      3.4.2 - Data Transformation ............................................................................................. 28
   3.5 - Data Partition ............................................................................................................... 30
   3.6 - Modeling ....................................................................................................................... 31
   3.7 - Evaluation and Results ................................................................................................. 31
      3.7.1 - ROC Curve ............................................................................................................ 31
      3.7.2 - Misclassification Rate ........................................................................................... 32
      3.7.3 - Cumulative Lift .................................................................................................... 33
      3.7.4 - Confusion Matrix ................................................................................................ 33
   3.8 - Deployment .................................................................................................................. 36

4. CONCLUSION ..................................................................................................................... 38

5. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS ................................. 39

6. BIBLIOGRAPHY .................................................................................................................. 40
LIST OF TABLES

Table 1 – Summary of related work on churn prediction in banking ................................. 16
Table 2 – Description of used variables ............................................................................. 24
Table 3 – Neural Network models ...................................................................................... 31
Table 4 – ROC curve results ............................................................................................... 32
Table 5 – Confusion Matrix results – Neural Network 1 .................................................... 34
Table 6 – Confusion Matrix results – Neural Network 2 .................................................... 34
Table 7 – Confusion Matrix results – Neural Network 3 & 4 ............................................... 34
Table 8 – Confusion Matrix metric results ......................................................................... 36
Table 9 – Confusion Matrix results – Score dataset .......................................................... 36
Table 10 – Confusion Matrix metric results – Score dataset ............................................. 37
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Categories of churners</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Age distribution of former churners</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Profession distribution of former churners</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Financial product distribution of former churners</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>District distribution of former churners</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>Single Layer Perceptron</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>Multi-Layer Perceptron</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>CRISP-DM Methodology</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>Import node - variable roles and levels</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>Stat Explore node - total number of observations</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>Multiplot node - Age</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>Stat Explore node - filtered number of observations</td>
<td>28</td>
</tr>
<tr>
<td>13</td>
<td>Replacement node</td>
<td>28</td>
</tr>
<tr>
<td>14</td>
<td>Variable selection</td>
<td>29</td>
</tr>
<tr>
<td>15</td>
<td>Variable worth</td>
<td>29</td>
</tr>
<tr>
<td>16</td>
<td>Data partition</td>
<td>30</td>
</tr>
<tr>
<td>17</td>
<td>ROC curve</td>
<td>31</td>
</tr>
<tr>
<td>18</td>
<td>Misclassification rate results</td>
<td>32</td>
</tr>
<tr>
<td>19</td>
<td>Cumulative lift</td>
<td>33</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 – BACKGROUND AND PROBLEM IDENTIFICATION

Following the debacle of the 2008 subprime financial crisis, the European Central bank (ECB) resorted to extreme accommodating monetary measures (such as quantitative easing and zero interest rate policies) in order to catalyze growth (An & Arch, 2014), rise inflation and lower unemployment rates. These measures, although beneficial to stimulate the economies of the breached countries, have been tremendously detrimental for the profitability of retail banks, whose principal source of income largely depends on the interest rate set by the ECB. This scenario has been ongoing for a number of years now, with no near term expected shift of policy from the ECB in sight. In an effort to counter these measures, banks were left with no choice but to find alternative sources of income, which in essence have consisted in containing costs and reviewing their pricing policies by charging higher commissions. Such measures, while successful in leveraging the complementary margins of banks (income gained through commissions), have had nonetheless a negative impact on customer satisfaction, and consequently have led to increased levels of customer churn.

Customer churn, which is defined as the propensity of customers to cease doing business with a company in a given time period, has become a significant problem and is one of the prime challenges many companies worldwide are having to face (Chandar, Laha, & Krishna, 2006).

The bank on which the present study will be conducted is currently one of the most well-known organizations in the Portuguese corporate sector, and accordingly is one of the most recognizable brands in the public eye. For long now, it has been a key driver of the development of the Portuguese economy by providing credit to consumers and businesses, and it was (up until recently) one of the pillars of the national financial system. Its long history of existence makes it possessive of one of the richest and most diverse client databases in the country. For confidentiality purposes, this organization will be referred to as bank XYZ in this study.

Over the last decade, the internet and the outburst of social media have had a profound effect on the retail banking business, granting customers easier and immediate access to information. The digitization of banking operations is ongoing and seems to be inevitable. As a result, bank XYZ (like its competitors) is currently undergoing a major restructuring of its branch network with overall downsizing ramifications in an effort to adjust to this new reality.

In what concerns day-to-day operations, the company currently makes use of its business intelligence structure to generate leads in accordance with pre-defined marketing campaigns. Regarding churn management, bank XYZ is aware by a recent marketing research study that
the overall increase in commissions have led to an escalation in the number of churners. Notwithstanding these findings, the company still only acts on this topic by engaging in marketing recovery campaigns, meaning it only addresses the issue of churn in a reactive way. In other words, it is only after a customer has churned that bank XYZ takes an active stance on the matter in an attempt to win back that same customer.

Marketing literature states that it is more costly to engage a new customer than to retain an existing loyal customer (Sharma & Kumar Panigrahi, 2011). Therefore banks now need to shift their attention from customer acquisition to customer retention, provide accurate churn prediction models, and effective churn prevention strategies as added customer retention solutions to preventing churn (Oyeniyi & Adeyemo, 2015).

1.2 – Study Objectives

This study proposes a new outline for dealing with customer churn in bank XYZ. Concretely, rather than addressing customer attrition in a reactive manner, a preemptive approach is suggested in order to tackle the issue from a proactive standpoint.

Customer churn prediction is the practice of assigning a churn probability to each customer in the company database, according to a predicted relationship between that customer's historical information and its future churning behavior. Practically, the probability to end the relationship with the company is then used to rank the customers from most to least likely to churn, and customers with the highest propensity to churn receive marketing retention campaigns (Coussement, Lessmann, & Verstraeten, 2016).

With this in mind, the overall goal of this study is to ultimately propose a new framework for churn management in bank XYZ, so that the company’s marketing efforts be also channeled towards the retention of its current customers rather than just the recovery of past churners. In order to accomplish this, the following objectives are to be carried out:

First, data mining techniques will be employed to study the behavior and patterns of a sample composed of past churners over a time period. In this step, it is our goal to uncover the most significant (independent) variables that ultimately had the greatest impact on the churning of those customers. Common churn management activity focuses on churn prediction using past churn data and the factors of customer churn known as predictors (Ismail, Awang, Rahman, & Makhtar, 2015).

The second stage of this process will be focused on using the variables from the first step as inputs to train a set of predictive models using neural networks, and check the performance of each model.
The third stage will entail the evaluation of the quality of the models by checking the performance of the classifier in each model. The success of churn prediction is determined by measuring the ability of the prediction models to produce high accuracy in correctly predicting whether a customer will churn or not (Ismail et al., 2015).

The next step of the process will consist on running the previously trained models with new data taken from a sample of current customers, and report the results obtained in each model according to the target variable (churner or non-churner).

Finally, the last stage will be focused on monitoring the behavior of those customers over a time period, and assessing the prediction quality of the models against the actual outcome of those customers.

1.3 – STUDY IMPORTANCE

The current study is relevant in the sense that it follows up on current work projects conducted by the company on the topic of churn. Recently, the bank conducted a study that involved the characterization (socio-demographic and socio-economic profiles) of a sample of former customers. Questionnaires were used to identify the most common reasons that caused the churning, whereby it was discovered that over a third of the churners left the company due to the disapproval with the overall increase of commissions. The aim of this study was to support the decision-making process regarding the development of new products and services that would increase the value and strengthen the relationship with its current customers. The realization of such study proves that the subject of churn is an issue that raises general concern for the company. In fact, churn will continue to exist and customer management is the best way to ensure sustainable business growth for long term profitability rather than capturing new customers (Ismail et al., 2015). It is our goal to lean on the results of these findings to assist in the choice of customers to use as sample for the present study.

Moreover, it should be noted that monitoring churn should be a constant concern of any company today, regardless of its industry, as competition is fierce and the digital age has made it easier for customers to move their business elsewhere if so is desired. It is, therefore, in the best interest of any company to keep track of the behavior of its customers in order to potentially anticipate any signs of dissatisfaction that could eventually lead to churning. Such actions may be instrumental to reach out to those customers and hopefully save the relationship with the bank. The underlying financial benefits of such efforts are straightforward, as potential churners chose to maintain their business with the bank. Correctly predicting that a customer is going to churn and then successfully convincing him to stay can substantially increase the revenue of a company, even if a churn prediction model produces a certain number of false positives (Mutanen, Ahola, & Nousiainen, 2006).
2. LITERATURE REVIEW

2.1 – DEFINITION OF CHURN

The concept of churn, also known as customer attrition, turnover, or defection, is hardly a novelty in any industry. As Oyeniyi and Adeyemo (2015) pointed out, “churning is an important problem that has been studied across several areas of interest, such as mobile and telephony, insurance, and healthcare. Other sectors where the customer churn problem has been analyzed includes online social network churn analysis, and the retail banking industries”. Although the broad or most generally accepted definition of churn refers to the loss of a customer by a specific company, one must analyze the concept with regards to the context in which it is being employed. Eichinger, Nauck, and Klawonn (2006) defined customer attrition when a customer is leaving for a competitor. This notion has been backed by Qiasi, Roozbehani, & Minaei-bidgoli (2002) who consider churn when a customer discontinues the use of an organization’s products and services in favor of a competitor’s products and services. On the other hand, Neslin et al. (2006) described customer churn as the propensity of customers to cease doing business with a company in a given time period.

In the banking industry, the scope of the term is wide and is currently being utilized within several different fields of the business. Credit card churn occurs when a customer ceases to use its credit card within a specific timeframe. Likewise, network banking churn may be defined as a customer who stops using its internet (home banking) service - Chiang, Wang, Lee, & Lin (2003) covered this topic by measuring the periodicity of transaction time of the users. Additionally, Glady et al. (2008) defined a churner as a customer with less than 2.500 Euros of assets at the bank (savings, securities, or other kinds of products), and therefore paved the way for two distinct definitions of churn that exist in the organization that will be studied in this paper: the notion of voluntary churn and involuntary churn.

The current study will be particularly geared towards tracking the behavioral history of past churners within a specific timeframe in order to pinpoint certain patterns that might indicate that a customer is in risk of churning. To collect the data regarding past churners, it is important to first shed some light on what grounds a customer is considered a churner, as the company does not erase a client record even if the customer ceases the business relationship (voluntary churn). Indeed, “at any time, customers can stop operating their accounts with the bank, and become a churner without leaving immediate trace. This implies that churn in such cases happens with no tracking point such as or inactive account this makes it difficult to to recognize churners” (Oyeniyi & Adeyemo, 2015). In this particular bank, a customer is considered an involuntary churner whenever the relationship is deemed of weak involvement. Weak involvement is a status defined when the following conditions are attained:
- The customer is not the first account holder of any current account
- There is no transaction record in the last three months in any current account of the customer
- The balance of any current account owned by the customer is lower than 50€, with no credit loans, credit cards, or financial assets associated.

Taking this information into account, it is important to determine the type of customers that will be considered in this project. The aforementioned conditions of weak involvement immediately pose some difficult challenges to overcome with regards to tracking involuntary churn. The fact that a customer is considered a churner whenever his balance drops below 50€ implies that there could be daily shifts between being a churner and non-churner. These sorts of situations will not be addressed in the present study, hence voluntary churners will be the main target focus of this project.

### 2.2 – Overview of the Retail Banking Industry and Effects on Churn

The following section aims to deliver a brief outlook on the current state of the retail banking industry in Portugal, as well as provide insight into future events that will affect this field and subsequently have an impact on customer churn in years to come.

Ten years after the financial crisis, the banking industry in Portugal has begun and is still undergoing a major restructuring of its operating structures. The move towards digital in response to market demand has resulted in the closure of hundreds of branches throughout the country, with the inevitable layoff of thousands of workers. These events have had expected repercussions on customer satisfaction, with fewer employees available to meet customer needs, and therefore having an impact on customer churn.

On a brighter note, the ECB expansionary measures have been essential in providing liquidity to the financial organizations of the EU, which in turn have resulted in increased lending to consumers and corporations all across the different EU nations. The effects of such measures have been overwhelming positive to economies across Europe by boosting employment and overall standard of living. However, it is important to point out that such measures have reignited pricing competition between financial organizations to win over credit operations, which ultimately have an effect on churn as customers are less reluctant to move their business elsewhere if presented with better pricing proposals.
In addition, banks have been countering the effects of lower financial margins by increasing commissions across the various fields of the business, which directly impacts customer satisfaction and thus churn levels. According to a recent questionnaire conducted by the bank that consisted in contacting past churners to understand the underlying reasons for them ceasing their business with the company, results revealed that churners usually fall into five different categories:

- 37% left the company due to the increased commissioning
- 19% ceased their business due to only wanting to work with one single financial organization
- 17% left the bank due to life circumstances (unemployment, family, or work-related issues)
- 15% ceased their business due to being displeased with the level of customer service
- 13% left the bank for better business proposals (lower credit rates, higher deposit rates, etc.)

The results of this questionnaire depict various reasons for customer churn, although it is clear that higher commissions are having the biggest toll on churn with over a third of customers leaving the company for this reason. Several factors can be accounted to explain such outcome. As previously mentioned, banks had to increase their complementary margins to
compensate for the decline of their financial margins, with fees rising as much as 60%. On the other hand, the financial crisis had profound negative effects on the Portuguese people’s perception of the safety, transparency, and trust of financial organizations, with two of the main retail banks in the country collapsing, leaving thousands of people without their hard-earned life savings. These events ended up generating an overall sentiment of suspicion and distrust towards financial organizations, which ultimately affected customer complacency levels towards increases in fees and other commissions. Still, it should be mentioned that market research studies show that bank XYZ has currently the lowest commissioning pricing out of all its main competitors. Taking into consideration that banking products are fundamentally similar between retail banks operating in the country, it is reasonable to assume there is a gap between the quality of service provided by the bank, and the level of service perceived by customers. The fact that customers are leaving bank XYZ for other competitors in spite of higher commissions in other banks should be a warning sign that operating procedures ought to be up for review.

Lastly, it is expected that the year 2018 will change the competitive landscape of retail banking, as the Revised Payment Service Directive (PSD2) will open the door for new players to enter the market. PSD2 will essentially enable bank customers, both consumers and businesses, to use third-party providers to manage their finances, meaning banks will be obligated to provide access to these third-party to their customers’ accounts, so banks will no longer be competing against banks, but against everyone offering financial services. As customers are becoming more digital and mobile in their approach towards companies, consumer trust and preference will tend to increase in favor of these new financial services companies. Traditional banks will need to ramp up their pace of innovation in order to secure their customers and keep their place in the market.

2.3 – A STUDY OF PAST CHURNERS

Having established the various categories of churners that affect this particular financial institution, it is important to dig deeper and actually look into the characteristics of past churners. Such information will prove to be valuable in the data collection process, in particular to extract the score dataset composed of actual customers which are more likely to churn. For this purpose, a dataset composed of former customers who churned during the year of 2016 was extracted from the company’s Data Warehouse, whereby several findings were drawn from this study.

Looking at Figure 2, it is apparent that most churning occurred in customers within the age bracket of [18-39] (45%), and of over 55 years of age (35%). The Profession figure indicates
that churning is most prominent within students (23%), followed by non-qualified workers (15%), and remaining professions evenly distributed between 4-9%.

**Figure 2: Age distribution of former churners**

```
<table>
<thead>
<tr>
<th>Age Group</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Until 17</td>
<td>4%</td>
</tr>
<tr>
<td>18-39</td>
<td>45%</td>
</tr>
<tr>
<td>40-54</td>
<td>15%</td>
</tr>
<tr>
<td>&gt; 55</td>
<td>35%</td>
</tr>
</tbody>
</table>
```

**Figure 3: Profession distribution of former churners**

- Students: 22.9%
- Non Qualified Workers: 15.3%
- Heavy Machinery Operators: 5.4%
- Qualified Workers: 9.0%
- Agriculture and Fishing Workers: 4.5%
- Service and Sales Staff: 6.0%
- Bank Employees: 0.5%
- Administrative Staff: 6.9%
- Intermediate Level Technicians: 4.6%
- Specialists of Intellectual and Scientific Professions: 6.9%
- Army Personnel: 0.5%
- Senior Public Administration Staff: 4.7%
- Unknown Profession: 13.0%
The following figure presents the various bank products held by churners in the year 2016 prior to churning. Debit cards and homebanking are the two most common products requested by new customers after opening an account, as they provide means to access and withdraw funds. It is, therefore, not surprising to find that these two financial products were the most frequently owned by churners (debit cards – 39,3%; homebanking – 29%). It is, also, understandable to find saving deposits ranked high on the list (15,3%), as a considerable number of customers tend to open a saving accounts after opening their current account.

Wage credit, i.e an order given by the customer to its employer to transfer its wage to a designated bank account, constitutes an unexpected finding within the financial products mostly held by churners (16,9%). Wage credit is an extremely important financial product for a bank, as it represents a sign of preference over other financial institutions. A customer may hold several accounts in different financial institutions for different purposes, nevertheless the account where a customer collects its pay check usually means that it is an account of reference for that customer. Moreover, wage credits, by representing incoming funds, make an important contribution to banks for solvency purposes. For these reasons, it is alarming to find wage credits among the financial products mostly held by former churners, as it signals that the bank is losing valuable customers from a profitability standpoint.

The remaining list of bank products are within the least owned by former churners. Their ownership by customers usually signals a high degree of involvement with the bank (mortgage credit – 1,5%; consumer credit – 0,4%), or are more complex from a financial literacy standpoint (fixed-term deposits – 4,4%; overdraft limit – 4,2%; credit cards – 3,7%; financial insurance – 2,9%; transferable securities – 1,5%; structured deposits – 0,4%).
Finally, the following figure presents the geographic spread of churners during the year 2016. It is immediately striking that the bulk of churners are from the Lisbon (22%) and Porto (16%) districts, with an even spread (1 to 8%) among all remaining Districts.
The issue of customer churn has always been relevant to most companies, although the increasing importance of data analytics to decision making, coupled with a vast array of Business Intelligence tools available in the market today, has resulted in banks taking advantage of such technologic advances to address the churning of its customers. Indeed, there is a substantial body of work on churn prediction models. Logistic regression, decision trees, neural networks, support vector machines and survival analysis are the most popular methods e.g. (Buckinx & Van den Poel, 2005; Coussement & Van den Poel, 2008; Karahoca & Karahoca, 2011; Neslin, Gupta, Kamakura, Junxiang, & Mason, 2006).

Handling churn in the banking industry is hardly a straightforward process as multiple techniques have been put in place to carry out such purpose in different fields of the business.
Chiang et al. (2003) opted for Association Rules to study the sequential patterns that customers follow before ceasing to use network banking services. Time windows and Normalization were important criteria that needed to be defined in order to find certain customer patterns (the smaller the time window, the lesser related rules were to be found).

Zhao, Li, Li, Liu, & Ren (2005) used support vector machines to address the wireless industry customer churners, and appraises the performance of this classifier against other more traditional algorithms such as Artificial Neural Networks, Decision tree, and Naive Bayes. Their opinion was that SVM “hold high potential against traditional approaches due to their scalability, faster training and running times”.

Mutanen et al. (2006) employed the logistic regression algorithm to predict customer churn in a Finish bank. For this purpose, a sample of 151,000 customers was used, carrying 75 variables from distinct categories, such as account transactions IN, account transactions OUT, service indicators, personal profile information, and customer level combined information. As the sample used contained only 0.5% of churners, the issue of class imbalance was addressed by under sampling the majority class (non-churners), thereby producing two different datasets (one with a churner/non-churner ratio of 1/1, and the other with a 2/3 ratio). Six regression models were produced by running the two datasets over three different time periods, and a lift curve was used to visualize the accuracy of prediction by each of the models, whereby five of the models had a prediction accuracy of over 60%.

Xie, Li, Ngai, & Ying (2009) proposed improved balanced random forests (IRBF) by combining balanced random forests (by oversampling the training data with the minority class) and weighted random forests (by assigning a larger weight to the minority class, and thus penalizing the misclassification of the minority class more heavily). The authors used a database of 20,000 bank customers containing 15 variables after removing descriptors that had too many missing variables. The remaining variables were grouped into three categories, namely personal demographics (age, education, size of disposable income, employment type, marital status, number of dependents, and service grade), account level (account type, guarantee type, length of maturity of loan, loan data, and loan amount), and customer behavior (account status, credit status, and the number of times the terms of agreement had been broken). An unbalanced training dataset was used as it “contains a proportion of churners that is representative of the true population to approximate the predictive performance in a real-life situation”. In the end, the authors used the lift curve to show that IRBF achieved a better performance against other classifiers, such as artificial neural networks, decision trees, and CWC-SVM.

Jinbo, Xiu, & Wenhuang (2007) ran three different variations of the Adaboost (Real, Modest, and Gentle) algorithm on a credit debt customer database of an anonymous bank in China. Adaboost is a classification algorithm that works by running “re-weighted versions of the training data and then taking a weighted majority vote of the sequence of classifiers thus
produced”. During the pre-processing stage, the authors opted to exclude all observations that had more than 30% of missing values, and chose to retain 19 variables composed of customer demographics, behavioral, and company interaction data. Missing values were treated according to the type of variable of study (continuous or categorical). The mean of the non-missing observations was used for continuous variables, and an indication of missing was introduced for categorical variables. After running the three different algorithms on balanced dataset, the authors found that the algorithm with “higher learning speed, less error rate or higher lift value is usually more overfitting”, whereas the slowest learning algorithm (Modest Adaboost) “resists the overfitting problem very well”. In the end, the study not only revealed that Adaboost algorithms could predict customer churn with a high likelihood, but also indicate the potential rules of the classification process by determining the attribute which had the higher influencing factor to the classification.

Kumar & Ravi (2008) developed an Ensemble system by majority voting using Multilayer Perceptron (MLP), Logistic Regression, decision trees, Random Forrest, Radial Basis Function network, and Support Vector Machines. A Classification and Regression Tree (CART) was employed for the purpose of feature selection. The issue of unbalanced data was dealt with several techniques – undersampling, oversampling, a combination of undersampling and oversampling, and by the Synthetic Minority Oversampling Technique (SMOTE). Results revealed that SMOTE achieved overall good accuracy.

Wang et al. (2010) conducted a study on credit card churn by going as far as to use twelve different algorithms (Simple Cart, J48, RandomTree, Logistic regression, SMO, Bayes networks, Naïve Bayes, KStar, IBK, decision table and PART). Their thesis relied on the assumption that “there exists no single algorithm that could achieve the best performance for all measures in a given application”, hence “the algorithm selection is a critical problem in the field of artificial intelligence and machine learning”. The different algorithms were assessed according to several evaluating criteria, such as Accuracy, AUC, Precision, Recall, Mean Absolute Error, and the Training and Testing time, and then ranked by employing two Multi-criteria decision-making processes (PROMETHEE II and TOPSIS). PROMETHEE II ranked alternatives according to their net flow (the higher the net flow, the better the alternative), and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) evaluated alternatives according to the distance between alternatives and the ideal solution. A sample dataset was extracted from the bank under study’s Data Warehouse containing a series of different customer attributes, namely: demographic variables, frequent variables, the total amount variables, extreme variables, and the status variables. Preprocessing work was conducted on this dataset in order to train the different models, and evaluation of the classifiers was later performed by using the two aforementioned methods. Results showed that the Logistic Regression and J48 Decision Tree algorithm were the best overall methods to employ to perform credit card churn analysis, albeit revealing slight differences between the two methods used, as TOPSIS ranked the Bayes Classifier as the best method.
Nie et al. (2011) used the Logistic Regression and Decision Tree algorithms to predict credit card churn in a Chinese bank. For this study, the authors defined two-time window periods – the observation period (12 months), whereby the transaction behavior of the customers was tracked, and the performance period (12 months) to track whether customers later became churners or not. To conduct this project, 135 variables were used, grouped into four different categories (customer personal information, card basic information, risk information, and card information). Afterwards, multicollinearity was employed to reduce the dimension space to 95 variables, and a balanced training set was used to train the models. Results showed that Logistic Regression performed better than the Decision Tree, with the card and transaction information being more significant for churn prediction than the customer demographic information.

Kaur (2013) employed the Naïve Bayes, Decision Tree (J48), and Support Vector Machine classifiers on a dataset of 2.000 customers containing both churned and active customers. The data was split between 70% for the training set, and the remaining 30% for the validation set. In the end, evaluation on all models (performed through a confusion matrix) disclosed a higher prediction rate for the churn class than for the loyal class, although the Decision Tree algorithm revealed a better prediction accuracy than the Naïve Bayes and Support Vector Machine.

Gür Ali & Aritürk (2014) deviated from the standard classifier approaches which considered only one customer observation at a certain point in time, and proposed a new framework entitled dynamic churn prediction which included customer observations from different time periods. This study was conducted on a sample of 7,204 private banking customers on a 12-month period, with the pre-processing stage involving the normalization of variables to mean 0 and standard deviation 1, and outlier elimination. A total of 169 variables grouped into four main categories were used, namely: customer behavior (current and historical portfolio value, return and product usage information), customer demographics (age, gender, education level, and nationality), customer-company interactions (service quality and timeliness of services), and economic indicators (consumer confidence index, consumer price index, local stock market index, USD exchange rate, and the yield curve slope). In the end, the authors showed that using “multiple observations training observations per customer from different time periods (MPTD) increases the prediction accuracy of churn models, compared with the traditional approach of using the most recent observation per customer, regardless of the classifier or prediction horizon”. In addition, the proposed framework revealed superior prediction accuracy against survival analysis (Cox regression).

Lastly, Zoric (2016) conducted a study on a Croatian bank using Neural Networks. A sample of 1,866 customers was extracted from the company’s database, with the most relevant variables being sex, age, private status (pensioners, employed, students, and unemployed), average monthly income, usage of internet banking and usage of two or more banking products (currency account, credit, savings, internet banking, mobile banking, SMS, standing
orders, etc.). Results showed that students were the most problematic group of customers in risk of churning (mainly due to holding a very low number of banking products), while pensioners were the group of customers with the lowest chances of churning. The bank under study should therefore address the student group by offering products tailored to their needs. It is worth mentioning that a change in the topology of the neural network did not affect the results of the study.
<table>
<thead>
<tr>
<th>Table 1: Summary of related work on churn prediction in banking</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain</strong></td>
</tr>
<tr>
<td>Finance</td>
</tr>
<tr>
<td>Marketing</td>
</tr>
<tr>
<td>Data Science</td>
</tr>
<tr>
<td>Economies</td>
</tr>
<tr>
<td>Customer Service</td>
</tr>
<tr>
<td>Operations</td>
</tr>
</tbody>
</table>
2.5 – ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANNs) are computational modeling tools that have emerged and found extensive acceptance in many disciplines for modeling complex real-world problems. As the name suggests, neural networks are brain-inspired systems intended to replicate the way humans learn. ANNs may be defined as structures comprised of densely interconnected adaptive simple processing elements (called artificial neurons or nodes) that are capable of performing massively parallel computations for data processing and knowledge representation (Hecht-Nielsen, 1990; Schalkoff, 1997). A neuron is a simple mathematical function capturing and organizing information according to an architecture. A neural network consists of layers of nodes connected by synapses, which are really weighted values.

2.5.1 – SINGLE LAYER PERCEPTRON

A single layer perceptron is the simplest form of a neural network as it contains only two layers of nodes – the input layer and the output layer. It is a feed-forward network based on a threshold transfer function that can only classify linearly separable cases with a binary target (0,1).

![Figure 6: Single Layer Perceptron](image)

The algorithm works by assigning random weights to the input values, and then summing the weighted values against a threshold ($\theta$).
If $W_1X_1+W_2X_2+W_3X_3+\ldots+\ldots+W_nX_n > \theta$

\[\rightarrow 1\]

If $W_1X_1+W_2X_2+W_3X_3+\ldots+\ldots+W_nX_n \leq \theta$

\[\rightarrow 0\]

In this step, an offset named bias is also added to the sum.

The performance of the classifier is considered satisfactory and entails no changes to the weights if the predicted output is close to the desired output. However, if there is a significant gap between the predicted output and the desired output, the neural network goes through a learning phase where it adjusts the weights based on the last obtained error. Weights and biases are gradually shifted so that the next result is closer to the desired output.

2.5.2 – **Multi-Layer Perceptron**

A multi-layer perceptron (MLP) has the same structure as a single layer perceptron with one or more hidden layers. In the case of an MLP, the input layer receives the input values, and transfers that information to one (or more) hidden layers, which are really a collection of nodes responsible for processing that information. The nodes are connected by weights and output signals which are a function of the sum of the inputs to the node modified by a simple nonlinear transfer, or activation function (Gardner & Dorling, 1998). The output of a node is scaled by the connecting weight and fed forward to be an input to the nodes in the next layer in the network (Rutkowski et al., 2006). By selecting a suitable set of connecting weights and transfer functions, it has been shown that a multilayer perceptron can approximate any smooth, measurable function between the input and output vectors (Hayati & Mohebi, 2007). During the training stage, the resulted output of the MLP may not equal the desired output. The gap between the desired output and the actual output is called the error signal. Training uses the magnitude of this error signal to determine to what degree the weights in the network should be adjusted so that the overall error of the multilayer perceptron is reduced (Gardner & Dorling, 1998). Hidden layers are therefore an intermediate layer between the input layer and the output layer, and their job is to act as the computational engine on the MLP by adjusting the weights and biases of the model in order to minimize the error.
2.5.3 – THE BACKPROPAGATION ALGORITHM

The backpropagation algorithm (Rumelhart et al., 1986) is the most computationally straightforward algorithm for training the multilayer perceptron. As previously mentioned, the goal of training is to find the best combination of weights in order to minimize the error against the actual desired output. The procedure used to accomplish this purpose is known as the gradient descent technique, whereby different weights are tested to find the minimum point in the error surface. It should be mentioned that the error surface may contain more than one minimum point named local minimums, which may not necessarily be the lowest minimum point of the error surface (named global minimum). Two parameters can be used to avoid the algorithm from being trapped in a local minimum (by assuming it is the global minimum): the learning rate and the momentum term. The learning rate determines the gradient descent curve and should not be too steep (therefore causing large weight changes which may result in the algorithm “jumping over” the global minimum), nor too small (causing the training to take too long). The momentum term is employed when the gradient descent gets stuck in a local minimum. It is a “boost” given to the current weights for the algorithm to leave the local minimum and keep searching for the global minimum.

In essence, the backpropagation algorithm can be summarized as follows:

- Randomly initialize the weights of the input vector of the training data to the network
- Propagate the input vector through the network to obtain an output
• Assess the error of the actual output against the desired output
• Propagate the error signal back through the network (hence the term backpropagation)
• Adjust the weights to minimize the error and repeat the process with the next input vector until the actual output is the closest to the desired output.
3. METHODOLOGY

The current project was carried out by making use of the most standard methodology for data mining purposes – the Cross-Industry Standard Practice for Data Mining, widely known as the CRISP-DM method.

![Figure 8: CRISP-DM Methodology](image)

The CRISP-DM methodology is a multinational, standards-based approach to describe, document, and continuously improve data mining (and associated data warehousing and business intelligence) processes (Ville, 2001), and is composed of six phases, namely:
3.1 – **BUSINESS UNDERSTANDING**

The goal of this project is to develop a predictive model with the use of neural networks able to estimate with a high degree of accuracy which customers are in risk of churning in a near future. To accomplish this task, the behavior of recent churners will be analyzed in the six-month period prior to churning, in order to uncover the most significant patterns which may be conducive to churn. A six-month window prior to churning is chosen as it is considered to be a reasonable timeframe to assess customer behavior that may lead to churn. Furthermore, as the goal of this project is to assess voluntary churn and not involuntary churn, an equivalent six-month timeframe is used to attest the duration of churn of the customers under study.

3.2 – **DATA UNDERSTANDING**

In order to track behavior, the following financial products were considered to undertake this study:

- **Debit Card**
  A payment card used to make cash withdrawals or to draw money directly from the checking account when a purchase is made.

- **Credit Card**
  A payment card used to borrow money against a line of credit, also known as the card’s credit limit.

- **Deferred Debit Card**
  A payment card that provides additional funds in addition to the funds existent in the customer’s current account. All payments made by the card are debited in the current account with deferral on a defined date in the next month.

- **Prepaid Card**
  A payment card that is not linked to the checking account, but instead works by loading money directly in it. Usually ordered by sponsors for teenagers for safety concerns, or used for managing expenses more efficiently.
• **Life Insurance**
  An agreement between a customer and an insurance company. Under the terms of a life insurance contract, the insurance company promises to pay a certain sum to someone (a beneficiary) in case of death, in exchange for premium payments.

• **Financial Insurance products**
  Financial asset products that gain from capitalization (growth in value over time). Such products include traditional savings schemes or 401(k).

• **Overdraft**
  An overdraft is an extension of credit from a lending institution when an account reaches zero. An overdraft allows the individual to continue withdrawing money even if the account has no funds in it or not enough to cover the withdrawal. Basically, overdraft means that the bank allows customers to borrow a set amount of money.

• **Mortgage Loan**
  A loan taken by a customer to purchase a real estate where the property is usually given as collateral.

• **Fixed-Term deposits**
  Fixed-term deposits are generally short-term deposits with maturities ranging anywhere from a month to a few years. When a term deposit is purchased, the customer understands that the money can only be withdrawn after the term has ended or by giving a predetermined number of days notice.

• **Savings deposits**
  A savings account is an interest-bearing deposit account that provides a modest interest rate. Savings accounts generally are opened to keep money not intended for daily or regular expenses.

• **Structured deposits**
  A combination of a traditional savings account and a stock market investment

• **Homebanking**
  Homebanking is the practice of conducting banking transactions from home rather than at branch locations. Home banking generally refers to either banking over the telephone or on the internet (i.e. online banking).
Below is a description of the variables used:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPVAR</td>
<td>Target variable indicating the existence of churn</td>
</tr>
<tr>
<td>I_CANC_ACT_CCRED</td>
<td>Flag variable indicating the credit card of the current account has not been utilized in the past three months</td>
</tr>
<tr>
<td>I_CANC_ACT_CDDIF</td>
<td>Flag variable indicating the deferred debit card of the current account has not been utilized in the past three months</td>
</tr>
<tr>
<td>I_CANC_ACT_CDEB</td>
<td>Flag variable indicating the debit card of the current account has not been utilized in the past three months</td>
</tr>
<tr>
<td>I_CANC_ACT_CXO</td>
<td>Flag variable indicating that the customer has not accessed its homebanking service in the past three months</td>
</tr>
<tr>
<td>I_CANC_ACT_PEPAG</td>
<td>Flag variable indicating the prepaid card of the current account has not been utilized in the past three months</td>
</tr>
<tr>
<td>I_CANC_CC</td>
<td>Flag variable indicating the cancellation of a credit card</td>
</tr>
<tr>
<td>I_CANC_CDEB</td>
<td>Flag variable indicating the cancellation of a debit card</td>
</tr>
<tr>
<td>I_CANC_CHAB</td>
<td>Flag variable indicating that the customer has ceased having a mortgage credit</td>
</tr>
<tr>
<td>I_CANC_CXO</td>
<td>Flag variable indicating the cancellation of the homebanking service</td>
</tr>
<tr>
<td>I_CANC_DDF</td>
<td>Flag variable indicating the cancellation of a deferred debit card</td>
</tr>
<tr>
<td>I_CANC_DEST</td>
<td>Flag variable indicating the cancellation of all structured products</td>
</tr>
<tr>
<td>I_CANC_DP</td>
<td>Flag variable indicating the cancellation of all fixed-term deposits</td>
</tr>
<tr>
<td>I_CANC_DPOUP</td>
<td>Flag variable indicating the cancellation of all savings accounts</td>
</tr>
<tr>
<td>I_CANC_LDN</td>
<td>Flag variable indicating the cancellation of the overdraft limit of the current account</td>
</tr>
<tr>
<td>I_CANC_PEPAG</td>
<td>Flag variable indicating the cancellation of a prepaid card</td>
</tr>
<tr>
<td>I_CANC_SEP</td>
<td>Flag variable indicating the cancellation of all financial insurance products</td>
</tr>
<tr>
<td>I_CANC_SEGVIDA</td>
<td>Flag variable indicating the cancellation of life insurance</td>
</tr>
<tr>
<td>I_CANC_VNCT</td>
<td>Flag variable indicating that the customer’s wage is not being credited in its current account</td>
</tr>
<tr>
<td>I_GNRO_CLI</td>
<td>Customer gender</td>
</tr>
<tr>
<td>N_CLIENTE</td>
<td>Customer ID</td>
</tr>
<tr>
<td>Q_ANO_CLI</td>
<td>Customer Age</td>
</tr>
</tbody>
</table>

Table 2: Description of Used Variables

A sample composed of 1,588 customers from the period of January to June 2017, which would become churners from the period of July to December 2017, was extracted from the bank’s Data Warehouse with the use of SAS Base. As previously mentioned, the behavior of customers prior to churning will be the main focus of study of this project. Behavior will be tracked by making use of binary variables to signal the various actions/inactions taken by customers which may indicate the chance of churning. For instance, if a customer held a credit card in January 2017 but ceased to hold it by June 2017, a flag variable of 1 will indicate that the customer cancelled its credit card during the period of February to June 2017.
Consequently, a flag variable of 0 will indicate that the customer either did not cancel its credit card during the period of January to June 2017, or never owned this financial product during this particular timeframe.

In addition, the final dataset used to train the model was also fed with a sample of actual customers by the month of December 2017 that presented similar behavior to those churners in the period of January to June 2017. As previously shown in the study of past churners during the year of 2016, students made up for nearly one quarter of the total number of churners during that year. Such statistic could be explained by their low level of involvement with the bank, as students usually only carry the most basic financial products (such as one current account and one debit card), hence being more predisposed to churn. Due to their low binding ties with the bank, it is considerably challenging to predict churn in students. For this reason, a decision was made to exclude students from the pool of actual customers that make up this sample. As the scope of actual customers that fulfilled such conditions was broad (close to 500,000 customers), a decision was also made to narrow down the field by looking only into customers that owned the same or fewer number of banking products in June 2017 in comparison to January 2017, meaning their level of involvement with the bank either stabilized or decreased during this time period. Furthermore, a study of the number of banking products held by the 1,588 churners revealed that roughly 80% of the sample owned three or less products in June when compared to January, so a decision was made to only consider customers in these circumstances. In the end, this process resulted in a dataset of 88,883 actual customers which presented similar behavior to those customers that churned.

3.3 - DATA PREPARATION

After importing the complete dataset with the use of the file import node, the first step is to establish the roles and levels of all variables. The “N_CLIENTE” variable role is changed to ID, as it is the key that differentiates every single observation of the dataset. The Stat Explore node was used to confirm the successful import of the dataset, with 1,588 churners and 88,883 actual customers.


**Figure 9: Import node - Variable roles and levels**

<table>
<thead>
<tr>
<th>Name</th>
<th>Role</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depvar</td>
<td>Target</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_ACT_CCRE</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_ACT_CDDIF</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_ACT_CDEB</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_ACT_CXR</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_ACT_PREPAG</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_CC</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_CDEB</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_CHAB</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_CXR</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_DDF</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_DEST</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_DP</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_DFOUP</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_LDNP</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_PREPAG</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_SEGF</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_SEGVIDIA</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_CANC_VNCNT</td>
<td>Input</td>
<td>Binary</td>
</tr>
<tr>
<td>I_NRO_CLI</td>
<td>Input</td>
<td>Nominal</td>
</tr>
<tr>
<td>N_CLIENTE</td>
<td>ID</td>
<td>Nominal</td>
</tr>
<tr>
<td>Q_AVG_CLI</td>
<td>Input</td>
<td>Interval</td>
</tr>
</tbody>
</table>

**Figure 10: Stat Explore node - total number of observations**

Distribution of Class Target and Segment Variables
(maximum 500 observations printed)

<table>
<thead>
<tr>
<th>Data Role</th>
<th>Variable Name</th>
<th>Role</th>
<th>Level</th>
<th>Frequency Count</th>
<th>Frequency Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAIN</td>
<td>Depvar</td>
<td>TARGET</td>
<td>0</td>
<td>88883</td>
<td>98.2447</td>
</tr>
<tr>
<td>TRAIN</td>
<td>Depvar</td>
<td>TARGET</td>
<td>1</td>
<td>1500</td>
<td>1.7553</td>
</tr>
</tbody>
</table>
3.4 - DATA PRE-PROCESSING

3.4.1 - Data Cleansing

The next step consists on exploring the data to check for any missing values, errors, and outliers in all variables. The Multiplot node is used to check the distribution and frequency of observations per variable. The variable “Q_ANO_CLI” (age) signals the need to filter the existing dataset, with several observations indicating the presence of students (ages < 21), as well as the possible existence of deceased individuals (ages > 95) which have not been flagged as such in the company’s database.

![Multiplot node - Age](image)

Figure 11: Multiplot node - Age

After running the filter node, a total of 5,126 observations were eliminated from the initial dataset, thereby resulting in a dataset of 85,345 observations (83,782 customers and 1,563 churners).
Figure 12: Stat Explore node – filtered number of observations

3.4.2 - Data Transformation

The Replacement node was used to convert the “I_GNRO_CLI” (gender) variable role from nominal to binary with an end goal of enabling the use of this variable in the modelling process. As such, the male gender depicted with an “M” in the “I_GNRO_CLI” variable was replaced with a “1”, and the female gender depicted with an “F” was replaced with a “0”, resulting in the creation of a new variable designated as “REP_I_GNRO_CLI”.

Figure 13: Replacement node

Figure 14 depicts the different variables which will be used to train the different neural network models.
Having concluded the data pre-processing stage, it is important to check the worth of the variables which will be used to train the different neural network models. As seen in figure 15, “REP_I_GNRO_CLI”, “I_CANC_DP”, AND “I_CANC_DEST” are the variables which carry the greatest predictive power, whereas “I_CANC_ACT_CXD”, “I_CANC_ACT_CDEB”, and “I_CANC_CDEB” are the least impactful variables.
3.5 - Data Partition

Having concluded the data pre-processing stage, the final dataset is now ready to be used for the end goal of building a predictive model. The first stage of this process consists on partitioning the data to train the different neural network models that will be tested later in the project. A standard split of 70% training-30% validation was used, meaning 59.739 records were used to train the different models, and the remaining 25.606 records were used to validate the results obtained in the training stage.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Numeric Value</th>
<th>Formatted Value</th>
<th>Frequency</th>
<th>Percent</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depvar</td>
<td>0</td>
<td>0</td>
<td>83782</td>
<td>98.1686</td>
<td>Depvar</td>
</tr>
<tr>
<td>Depvar</td>
<td>1</td>
<td>1</td>
<td>1563</td>
<td>1.8314</td>
<td>Depvar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Numeric Value</th>
<th>Formatted Value</th>
<th>Frequency</th>
<th>Percent</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depvar</td>
<td>0</td>
<td>0</td>
<td>58646</td>
<td>98.1704</td>
<td>Depvar</td>
</tr>
<tr>
<td>Depvar</td>
<td>1</td>
<td>1</td>
<td>1093</td>
<td>1.8296</td>
<td>Depvar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Numeric Value</th>
<th>Formatted Value</th>
<th>Frequency</th>
<th>Percent</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depvar</td>
<td>0</td>
<td>0</td>
<td>25136</td>
<td>98.1645</td>
<td>Depvar</td>
</tr>
<tr>
<td>Depvar</td>
<td>1</td>
<td>1</td>
<td>470</td>
<td>1.8355</td>
<td>Depvar</td>
</tr>
</tbody>
</table>

Figure 16: Data partition
3.6 - MODELING

Four neural network models were attempted, with the sole differentiating factor between them being the number of hidden layers.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Hidden Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network 1</td>
<td>1</td>
</tr>
<tr>
<td>Neural Network 2</td>
<td>2</td>
</tr>
<tr>
<td>Neural Network 3</td>
<td>3</td>
</tr>
<tr>
<td>Neural Network 4</td>
<td>4</td>
</tr>
</tbody>
</table>

*Table 3: Neural Network models*

3.7 - EVALUATION AND RESULTS

3.7.1 - ROC Curve

The ROC (Receiver Operating Characteristic) curve is a graph showing the performance of a classification model at all classification thresholds. It is a plot of the true positive rate against the false positive rate (the relationship between sensitivity and specificity). An area under the curve of 0.8 means that a randomly selected case from the group being a churner (having a value of DepVar=1) has a score larger than that for a randomly selected case from the group who is not a churner (having a value of DepVar=0) in 80% of the time. When there is a perfect separation of the two groups (no overlapping of the distributions), the area under the curve reaches to 1. In other words, the closer the graph is to the top and left-hand borders, the more accurate the test (and the closer the graph is to the diagonal, the less accurate the test).

*Figure 17: ROC curve*
Looking into the model comparison fit statistics, all models demonstrate high predictive accuracy with ROC curve values higher than 0.8 in all datasets (train and validation).

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Train</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network (1)</td>
<td>0.837</td>
<td>0.829</td>
</tr>
<tr>
<td>Neural Network (2)</td>
<td>0.881</td>
<td>0.88</td>
</tr>
<tr>
<td>Neural Network (3)</td>
<td>0.892</td>
<td>0.883</td>
</tr>
<tr>
<td>Neural Network (4)</td>
<td>0.882</td>
<td>0.874</td>
</tr>
</tbody>
</table>

*Table 4: ROC curve results*

### 3.7.2 - Misclassification Rate

The Misclassification Rate is a measure of the error of the classifier, meaning the model classifies a churner as a non-churner, and a non-churner as a churner. Figure 19 shows that the training misclassification rate stops decreasing with a three hidden layer network, which means that Neural Network 3 achieves the best results in terms of prediction accuracy when compared to the remaining models - Neural Network, Neural Network 2, and Neural Network 4. In addition, the error of the classifier is very low for both the training and the validation set, and the difference in the error between both sets is not significant, which means that there is less chance that the model overfits the training data, and is performing well at classifying the class on new data (validation set).

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Valid:</th>
<th>Train:</th>
<th>Valid:</th>
<th>Train:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Misclassification Rate</td>
<td>Squared Error</td>
<td>Misclassification Rate</td>
<td>Squared Error</td>
</tr>
<tr>
<td>Neural Network (4)</td>
<td>0.010312</td>
<td>0.010952</td>
<td>0.011818</td>
<td>0.010248</td>
</tr>
<tr>
<td>Neural Network (3)</td>
<td>0.010580</td>
<td>0.010779</td>
<td>0.011861</td>
<td>0.010997</td>
</tr>
<tr>
<td>Neural Network (2)</td>
<td>0.010580</td>
<td>0.010355</td>
<td>0.011935</td>
<td>0.010225</td>
</tr>
<tr>
<td>Neural Network</td>
<td>0.010800</td>
<td>0.010115</td>
<td>0.010296</td>
<td>0.010036</td>
</tr>
</tbody>
</table>

*Figure 18: Misclassification rate results*
3.7.3 - Cumulative Lift

Lift is a measure of the effectiveness of a predictive model calculated as the ratio between the results obtained with and without the predictive model. The Y-axis shows the percentage of positive responses (the percentage of total possible positive responses), and the X-axis shows the percentage of customers reached. A baseline (no model) is a horizontal line intersecting the Y-axis at 1, which means that if a random 10% of the dataset is reached, only 10% of the set will be customers at risk of churning. The chart of the validation set in figure 20 shows that by using the model, if only 10% of the customers are reached, a cumulative lift of 7 is obtained, which is 7 times as many if no model is used. That is the lift over the baseline, which means that the model is performing very well.

![Cumulative Lift Chart](image)

*Figure 19: Cumulative lift*

3.7.4 - Confusion Matrix

A Confusion Matrix contains information about actual and predicted classifications done by a classification system. Performance of such a system is commonly evaluated using the data in the matrix (Polat, Güneş, & Arslan, 2008). The Confusion Matrix has four categories: True positives (TP) are examples correctly labeled as positives; False positives (FP) refer to negative examples incorrectly labeled as positive; True negatives (TN) correspond to negatives...
correctly labeled as negative; False negatives (FN) refer to positive examples incorrectly labeled as negative (Davis & Goadrich, 2006).

Table 5: Confusion Matrix results – Neural Network 1

<table>
<thead>
<tr>
<th></th>
<th>actual positive</th>
<th>actual negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicted positive</td>
<td>TP = 0</td>
<td>FP = 0</td>
</tr>
<tr>
<td>predicted negative</td>
<td>FN = 470</td>
<td>TN = 25.136</td>
</tr>
</tbody>
</table>

Table 6: Confusion Matrix results – Neural Network 2

<table>
<thead>
<tr>
<th></th>
<th>actual positive</th>
<th>actual negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicted positive</td>
<td>TP = 243</td>
<td>FP = 69</td>
</tr>
<tr>
<td>predicted negative</td>
<td>FN = 227</td>
<td>TN = 25.067</td>
</tr>
</tbody>
</table>

Table 7: Confusion Matrix results – Neural Networks 3 & 4

<table>
<thead>
<tr>
<th></th>
<th>actual positive</th>
<th>actual negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicted positive</td>
<td>TP = 241</td>
<td>FP = 64</td>
</tr>
<tr>
<td>predicted negative</td>
<td>FN = 229</td>
<td>TN = 25.072</td>
</tr>
</tbody>
</table>
Based on the Confusion Matrix results, several metrics can be used to evaluate the quality of the algorithms under study:

- **Accuracy (AC)** is defined as the proportion of the total number of predictions that were correct.
  \[
  AC = \frac{TP + TN}{TP + TN + FP + FN}
  \]

- **Sensitivity**, or **True Positive Rate (TPR)**, is the proportion of positive cases that were correctly identified.
  \[
  TPR = \frac{TP}{TP + FN}
  \]

- **Specificity**, or **True Negative Rate (TNR)**, is the proportion of negative cases that were classified correctly.
  \[
  TNR = \frac{TN}{TN + FP}
  \]

- **The False Positive Rate (FPR)** is the proportion of negative cases that were incorrectly classified as positive.
  \[
  FPR = \frac{FP}{FP + TN}
  \]

- **The False Negative Rate (FNR)** is the proportion of positive cases that were incorrectly classified as negative.
  \[
  FNR = \frac{FN}{FN + TP}
  \]

- **Precision (P)** is the proportion of the predicted positive cases that were correct.
  \[
  P = \frac{TP}{TP + FP}
  \]
The following figure displays the results obtained for all the aforementioned metrics related to the Confusion Matrix:

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>False Positive Rate</th>
<th>False Negative Rate</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Neural Network 2</td>
<td>0,9877</td>
<td>0,5120</td>
<td>0,9973</td>
<td>0,0027</td>
<td>0,4829</td>
<td>0,7788</td>
</tr>
<tr>
<td>Neural Network 3 &amp; 4</td>
<td>0,9878</td>
<td>0,5128</td>
<td>0,9974</td>
<td>0,0025</td>
<td>0,4872</td>
<td>0,7902</td>
</tr>
</tbody>
</table>

*Table 8: Confusion Matrix metric results*

It can be observed that Neural Network 3 & 4 achieve the best results with regards to Accuracy, Specificity (TNR), False Positive Rate, and Precision metrics, whereas Neural Network 2 scores slightly more favorably in the Sensitivity and False Negative Rate metrics.

### 3.8 - Deployment

The concept of Deployment in Data Mining refers to the application of a model on new data. For this purpose, a new dataset was extracted from the company’s Data Warehouse containing customers during the period of July-December 2017 at risk of churning during the following six-month period (January-June 2018). To obtain this dataset, the same guidelines applied in the Data Preparation stage were followed. In other words, customers were deemed in risk of churning by displaying shifts of behavior during the time period under study. In addition, the final dataset was restricted to customers who owned three or less banking products by the end of the December 2017, and carrying an equal or fewer number of banking products when compared to July 2017.

A total of 134,842 customers that fulfilled these conditions was obtained from this process. This dataset was imported onto the current project in SAS Miner with the purpose of being scored by the model which delivered the most encouraging results – Neural Network 3. The Confusion Matrix displayed below summarizes the results obtained:

<table>
<thead>
<tr>
<th></th>
<th>actual positive</th>
<th>actual negative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>predicted positive</strong></td>
<td>TP = 565</td>
<td>FP = 19,945</td>
</tr>
<tr>
<td><strong>predicted negative</strong></td>
<td>FN = 1,915</td>
<td>TN = 112,417</td>
</tr>
</tbody>
</table>

*Table 9: Confusion Matrix results – Score dataset*
The total number of churners from the period of January-June 2018 was 3,339, so the final dataset of 134,842 customers contains 74.3% (565+1.915/3.339) of the total number of customers that ended up churning during this period. These results are consistent when compared to the final dataset used to train the models, where roughly 80% of the customers ended up churning.

Taking into account the above-mentioned results, the same metrics used beforehand were calculated to attest the quality of the model, whereby the results are shown in Table 10.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>False Positive Rate</th>
<th>False Negative Rate</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network 3</td>
<td>0.8401</td>
<td>0.2278</td>
<td>0.8493</td>
<td>0.1507</td>
<td>0.7722</td>
<td>0.0275</td>
</tr>
</tbody>
</table>

*Table 10: Confusion Matrix metric results – Score dataset*

When presented with new data, Neural Network 3 reveals very encouraging results with regards to Accuracy, Specificity, and False Negative Rate metrics. Sensitivity and False Positive Rate also display reassuring results, meaning the model is scoring nearly 23% of the total number of churners correctly, and only inaccurately scoring a non-churner as a churner in 15% of the total customer dataset. Precision is the only metric that does not deliver satisfactory results, as the model scores correctly a very low number of actual churners (565) when compared to the total number of predicted churners.
4. CONCLUSION

Following an extended period of time with historic low interest rates, the banking industry in Portugal has been undertaking major structural reforms in its business model to meet profitability goals. On one hand, revenue has been boosted by the increasing and substantial rises in commissions in an effort to counter low financial margins. On the other hand, the aforementioned reforms have been driven mainly by a decision to move towards digital, which has resulted in the downsizing of operations and closing of several branches. Such decisions have had a negative impact on customer satisfaction, and inevitably increased churn rates.

This project intended to provide a basis to address churn in a bank that is currently not making use of its rich database and analytic tools to face this critical issue. The first step consisted in extracting a dataset of customers during the period of January-June 2017, which would become churners during the period of July-December 2017. The goal was to track the behavior of these customers during the period of January-June 2017, which would possibly indicate a risk of churn in the near future. As such, the selection of variables relied entirely on the use of dummy variables that reflected a decrease in the level of involvement with the bank, meaning customers owned fewer financial products in June in comparison to January. Pre-processing work was conducted with the purpose of outlier elimination and data transformation, to which it was concluded that the most significant variables to train the proposed predictive models were customer gender and savings products (namely fixed-term deposits and structured deposits). Four neural network models differentiated by the number of hidden layers were trained, each delivering encouraging results in all studied metrics, albeit the Neural Network 3 model scoring slightly more favorably than remaining models. Model assessment was performed on a new dataset of 134,842 customers during the period of July-December 2017 at risk of churning during the period of January-June 2018. Results revealed an accuracy of 84%, with nearly a quarter of churners predicted correctly.

In the end, this project served its purpose to present a reliable alternative to anticipate and monitor customer churn behavior as opposed to the current reactive approach undertaken by the bank under study, which consists on developing marketing strategies focused on regaining past customers who turned churners. In light of the encouraging results displayed in this work, the current methodology led throughout this project could prove to be a valuable tool to predict churn in a company that has yet to make full use of the Business Intelligence tools at its disposal to tackle this issue.
5. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

The primary goal of this project was to develop a neural network model able to predict with a high degree of accuracy those customers which were at risk of churning. Despite revealing encouraging results with regards to several of the proposed metrics in the model development stage, the input of new data to assess the quality of the finished model delivered a lower than expected Precision rate when considering every customer that churned in that timeframe. These results demonstrate that some revisions are in order with regards to the customer behavioral methodology that was followed to fulfill the purposes of this project. This work considered the results of a characterization study of all churners during the year of 2016, which proved useful but at the same time insufficient to significantly narrow down the ratio of churners/non-churners which displayed similar behavior. In order to reduce the pool of non-churners, a decision was made to only take into account those customers which had equal or fewer number bank products at the end of the six-month time window period considered for the customer behavioral study. This decision, taken under the rational standpoint that customers in risk of churning should display behavior that signals the beginning stages of detachment from the bank, limited the pool of customers that ended up churning, thus affecting the accuracy results in the model assessment stage. For future developments, it would be interesting to re-assess the behavioral study of past churners with an end goal of lowering the pool of non-churners in order to reduce the ratio of churners/non-churners.

Furthermore, this project used a real-life dataset composed of the actual ratio of churners/non-churners to train the proposed models. Indeed, although many authors emphasize the need for a balanced training sample in order to differentiate reliably between churners and non-churners (Dekimpe & Degraeve, 1997; Rust & Metters, 1996), a decision was made to use a training dataset which contains a proportion of churners that is representative of the true population to approximate the predictive performance in a real-life situation (Xie et al., 2009). Still, random sampling methods (over-sampling or under-sampling) are often used to predict churn in order to reduce class imbalance. Through the way of sampling, for one thing, the training sample size is reduced and the model training speed is improved; for another thing, the capability of identifying churners of the training model is much better due to the reduced gap between the number of customers and that of non-customers (He, Shi, Wan, & Zhao, 2014).

In addition, the current project focused solely on estimating churn by considering data from one single six-month time window. Hence, the results presented could prove to be biased when compared to datasets involving different time periods. In the end, the proposed methodology would be reinforced by a study involving datasets from different time periods to further validate the results presented in this work.

Finally, this project focused solely on developing a neural network approach to estimate churn, hence overlooking the potential benefit of considering and comparing alternative models such as Decision Trees, Logistic Regression, or Support Vector Machines.
6. BIBLIOGRAPHY


Kaur, M. (2013). Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers.


