Prediction of Customer Status in Corporate Banking Using Neural Networks

Stanislaw Osowski
Military University of Technology
Warsaw University of Technology, Warsaw, POLAND
stanislaw.osowski@ee.pw.edu.pl

Lukasz Sierenski
Warsaw University of Technology, Warsaw, POLAND
ls@ee.pw.edu.pl

Abstract*—This paper presents a computer system that is based on the application of artificial neural networks and support vector machine and used to predict the future status of corporate banking clients. Three different states of clients are predicted: active (i.e., fully involved in business with a non-zero balance in their account), non-active (i.e., the balance in their account is very low and they own no forms of investments) and churning (closed bank account). The main task is to predict the likelihood of the corporate banking client departing, based on their previous banking activity. The most important element in such a prediction system is the development of the diagnostic features, which will serve as the input attributes to the classifier and are responsible for final recognition of the status of the customer. We propose a set of input attributes based on the financial activity of the client in the last 6 months prior to the prediction point. Two different classifiers will be checked and compared: a multilayer perceptron and a support vector machine. The results of their application have shown very good system operation quality.

Keywords*— Decision support systems; neural networks; data mining; prediction.

I. INTRODUCTION

Corporate banking refers to the aspect of banking that deals with corporate customers. It typically serves a diverse range of clients, ranging from small or mid-sized companies with a few millions in revenue to large conglomerates with billions in sales and offices across the country. An important factor in bank policy is to predict the future behaviour of a client on the basis of their past activity. Three different states of clients are the most important: active, i.e., fully involved in business with non-zero balance in their account, non-active, which is characterized by very low balance in the account and no forms of investments, and closed, where the client closes the bank account. From the business perspective, the most important task is to predict the possibility of the corporate banking client departing (usually called churning) on the basis of an analysis of his previous banking activity. However, the typical steps leading to churning lead from the active through the non-active stage. The recognition of entry into the non-active stage enables the bank to take some actions in order to make the client recover their active status.

Customer retention is crucial in a variety of businesses, including banking. Losing clients means losing money, as acquiring new customers can cost much more than satisfying and retaining existing customers. By minimizing customer churn, a company can maximize its profits. Early prediction of so-called possible "churners" enables banks to undertake some preventive measures for their retention. As churn management is a major task for companies hoping to retain valuable customers, the ability to predict customer churn is critical. As a consequence, churn prediction has attracted considerable attention from the business as well as academic worlds.

In literature, neural networks have shown their high applicability in churn prediction. Recently, the topic of customer churn prediction has been discussed extensively in several domains such as telecommunications [8],[9],[19], retail markets [1], and financial services provided by banks [4],[6],[7],[12],[18]. Many data mining techniques have been successfully applied in customer churn prediction. These techniques include artificial neural networks (ANNs) [15],[17], decision trees [2],[13], logistic regression [1],[2],[13], naïve Bayesian methods [13], random forest [1],[10],[21], fuzzy logic [11], Support Vector Machine (SVM) [3],[7] and even deep learning [20]. In particular, SVM has been found to be a very reliable and well performing classifier, used in many practical applications. It has strong theoretical foundations [14] and is currently a state-of-the-art method in machine leaning.

Relatively less attention has been paid to banking applications, and most papers are related to retail banking. The application of artificial intelligence methods is used mainly in predicting the possibility of churning [4],[10],[11],[12]. Approaches for the prediction of customer behaviour in corporate banking are difficult to find in the literature.

This paper will present a neural network approach to classify the state of clients in corporate banking. The procedure comprises the following steps: collecting the proper set of data on the basis of past activity of multiple clients, developing the diagnostic features representing the input attributes to the classification system, and constructing classifiers that can efficiently recognize the class with which the client should be associated.

II. ANALYSIS OF DATABASE USED IN EXPERIMENTS

The prediction process will be performed for corporate banking. In contrast to retail banking, corporate banking typically serves a diverse range of clients ranging from small companies to large conglomerates, which run businesses in diverse fields.

Four segments of clients of the chosen Polish bank have been considered in this study. In the first segment, we included the companies having annual sales below 10 million. Companies having annual sales between 10 and 50 million were ranked into the second group. The third and fourth segments comprised companies having annual sales between 50 and 200 million and above 200 million, respectively. These selections of client segments were normally used by the bank.

The bank needs to track customers' data over time, in order to provide them proper service. Each customer is described through many different records within the database for tracking operations over time. Therefore, the data have to be aggregated in order to be analysed, and this leads to the additional preprocessing stages. In our case, we have aggregated the selected actions month-wise to make database operations more manageable.

According to bank experts, the most important factors describing corporate clients belong to two basic groups. One is the total bank profit within the considered period of time (in our case, one month). It is formed by interest rates from the loans and securities granted, profits from all financial operations, exchange of currency, and certain other factors. The second group represents the volume of balance products concerning the particular client. It is the sum of loans, factorings and leasing. Statistical analysis of different components forming these two groups within the assumed period of time have shown, that the most important, for our classification problem, is the average monthly volume of the following items of the client:

- total result of banking activity
- product balances (actual bank commitment resulting from its past activity, for example loans granted and leasing)
- transaction banking (e.g., monetary flow and commercial papers)
- bank credits (sum of loans and overdrafts)
- credit products (sum of balance and off-balance products)
- leasing
- factoring
- trade financing (e.g., guarantees and letter of credits)
- treasure (including interest bearing securities, currencies, and shares listed on the stock market)

We have collected the monthly aggregated data related to the period between January 2016 to July 2018. The collected data included 1500 records related to 1500 customers having different status (active, non-active and churned). They represented all four segments of clients. The distribution of customers of different segments was as following: segment 1 – 345, segment 2 – 598, segment 3 – 330 and segment 4 – 227. The prediction for client status “Active” or “Non-active” was made for the last considered month (July 2018). In the case of status “Closed,” it was the last month within an analysed period of two years, in which the client left the bank. In all cases, the data related to 6 preceding months have been used in preparation of learning and testing samples.

III. Generation of Diagnostic Features

One of the most important elements in building a good classification system is to define the proper diagnostic features, which serve as the input attributes to the classifier. These can be used to build the mathematical model of the analysed data, suitably representing the mechanism of decisions taken by the customers. Such features should belong to a similar range of values, not dependent on themselves and well correlated with the class. The features proposed in this work will be based on 9 volume data presented in the previous section. The introductory correlation analysis has shown that the decision of class recognition in the nth month is strongly related to the volume values in the preceding 6 months. Moreover, the membership in the proper segment to which the customer belongs is also important (i.e., one of four segments coded in a binary way).

In developing a proper set of features, we should consider the dynamic changes of the descriptor values, calculated from month to month. Denoting the numerical descriptor in the nth month by x(n) and its monthly increment by Dx, i.e.

$$Dx(n)=x(n)-x(n-1) \tag{1}$$

we get the additional set of descriptors defined in each monthly period for 6 preceding months. To smooth the rapid changes of descriptor values from month to month, the moving average, covering three neighbouring months, has been applied and is defined as follows

$$Dx(n) := \frac{Dx(n)+Dx(n-1)+Dx(n-2)}{3} \tag{2}$$

To deal with the very high values of some descriptors, the additional logarithmic transformation has been applied. In such cases, these features have been defined as follows

$$Logx = \begin{cases} \log_{10}(x) & if \quad x>0 \\ -\log_{10}|x| & if \quad x<0 \\ 0 & if \quad x=0 \end{cases} \tag{3}$$

All these forms of features have been determined for 6 preceding months, i.e., from n to n−6. They create the multi-month attributes. The data have been gathered in the form of matrix X, in which the rows represent the observations in succeeding months, and columns represent the diagnostic features. As a result, the set of diagnostic features used as the input attributes to the classification system is composed of the following items.

1. Binary code of the segment membership of the client. There are 4 segments, coded as follows:
   1 0 0 0 – segment 1
   0 1 0 0 – segment 2
   0 0 1 0 – segment 3
   0 0 0 1 – segment 4

The next features have been calculated separately for each of the 6 months preceding the month for which the prediction is made. They have been defined as follows.

2. Total average volume in logarithmic scale, calculated according to (3).

3. The monthly increment of the total average volume in logarithmic scale, calculated according to (2).
4. The total volume of product balance in logarithmic scale.
5. The monthly increment of the total volume of product balance in logarithmic scale.
6. The total volume of transaction banking coded as 1, when its value is above the assumed minimum threshold and 0 in the opposite case.
7. The total volume of transaction banking in logarithmic scale.
8. The monthly increment of the total volume of transaction banking in logarithmic scale.
9. The total volume of bank credits coded binary as 1, when its value is above the assumed minimum threshold and 0 in opposite case.
10. The total volume of bank credits in logarithmic scale.
11. The monthly increment of the total volume of bank credits in logarithmic scale.
12. The total volume of credit products coded in binary as 1, when its value is above the assumed minimum threshold, and 0 in the opposite case.
13. The total volume of credit products in logarithmic scale.
14. The monthly increment of the total volume of credit products in logarithmic scale.
15. The total volume of leasing products coded binary as 1, when its value is above the assumed minimum threshold and 0 in the opposite case.
16. The total volume of leasing products in logarithmic scale.
17. The monthly increment of the total volume of leasing products in logarithmic scale.
18. The total volume of factoring products coded binary as 1, when its value is above the assumed minimum threshold and 0 in the opposite case.
19. The total volume of factoring products in logarithmic scale.
20. The monthly increment of the total volume of factoring products in logarithmic scale.
21. The total volume of trade financing coded binary as 1, when its value is above the assumed minimum threshold and 0 in the opposite case.
22. The total volume of trade financing in logarithmic scale.
23. The monthly increment of the total volume of trade financing in logarithmic scale.
24. The total volume of treasure coded binary as 1, when its value is above the assumed minimum threshold and 0 in the opposite case.
25. The total volume of treasure in logarithmic scale.
26. The monthly increment of the total volume of treasure in logarithmic scale.
27. Code of activity status of client in the particular month: 1 – represents active client when moving average value of the result of banking activity is above the assumed minimum threshold and 0 in the opposite state.

In this way, the total number of features is 160. All these features have been normalized, dividing real value by their median for all observations. In further experiments, we will show how the reduction of the preceding months (and number of features) influences the quality of the prediction process.

From the point of view of the bank, the most important is prediction of the possibility of churning the customer. The features presented above, related to 6 preceding months, have been found as characterizing such case in an appropriate way. Fig. 1 depicts the values of the succeeding diagnostic features representing two types of clients: active (the upper plot) and churning (the lower plot). The lower plot suitably illustrates the vanishing activity of a client who finally left the bank.

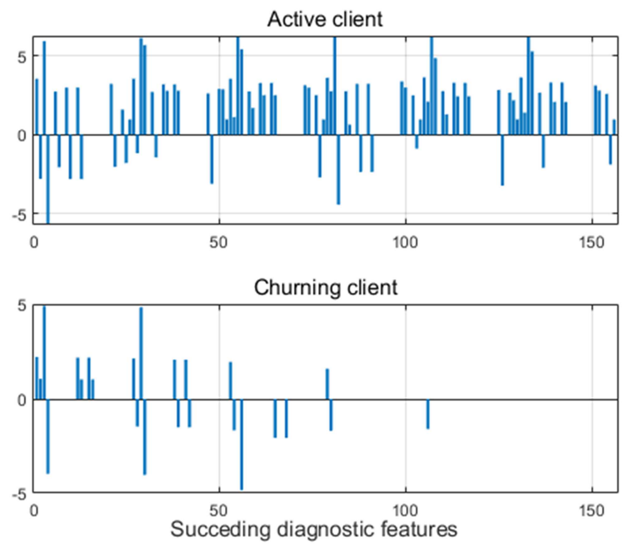

Fig. 1 Illustration of activity of two types of clients: active and churning.

IV. Classification Systems

Once the diagnostic features have been defined and generated on the basis of the available data, the next step is to create the most efficient classification system. According to the actual state of knowledge, the most efficient are the classifiers belonging to neural network family, especially multilayer perceptron (MLP) and support vector machine (SVM). These two classifier systems will be applied and compared in further investigations. We have checked also random forest of decision trees, however, its accuracy was a bit lower. Therefore, we have limited our presentation to only these two classifiers.

MLP network is one of the most universal neural approximators. It is the feedforward, multilayer artificial neural network structure, applying sigmoidal neurons. MLP consists of at least three layers of nodes: an input layer of input signals, a hidden layer of sigmoidal neurons and an output layer of sigmoidal neurons representing the classes. The number of output neurons is equal to the number of classes and each neuron represents a particular class. MLP applies a supervised learning technique based on gradient optimization, which is responsible for minimization of error function on the learning data set through adaptation of weights.

The SVM network of the Gaussian kernel, applied in solution, is one the most efficient classifiers currently [14]. The learning task of SVM is reduced to the quadratic optimization problem and is dependent on a few hyperparameters: the regularization constant C and the width parameter σ of the Gaussian kernel. Both are adjusted by repeating the learning experiments for the limited set of their predefined values and accepting this one, which results in minimum error on the validation data set. SVM has only one output neuron; hence, in the case of many classes, we train many 2-class recognizing networks applying the strategy either one class against one or one class against the rest. The final decision is taken on the majority voting principle. We have applied one against one strategy. For three recognized classes, three SVM networks have formed the classification system.

V. Results of Numerical Experiments

Numerical experiments have been performed on 1500 samples representing the data concerning 1500 corporate clients representing all 4 segments. They have been selected in a way that provides equal numbers of all three classes (each class represented by 500 observations). The observations, forming the input matrix **X**, have spanned a maximum of 6 preceding months prior to the actually predicted status of the client.

The important point in data preparation was very careful annotation of the classes. In the case of 1500 clients it was possible to do it manually by the bank expert. The second set of experiments was done on much larger set containing 10224 clients of the non-balanced ratio between classes. In this case the association of client with the class was done half-automatically, without personal intervention of the bank expert.

Two classification systems have been tested: the MLP network of 10 hidden sigmoidal neurons and 3 output neurons representing the three recognized classes and three SVM networks of Gaussian kernel (C=100, σ=1) working in one against one mode. These structures have been found as the best after the introductory experiments, in which 70% of the data has been used in the learning model and the other part used in verification of the learning process (testing data). The quality of model has been assessed on the basis of results on the testing data set.

To find the influence of the number of preceding months used in input data preparation, we have also conducted experiments with a reduced number. The MLP structure containing 10 hidden and 3 output neurons at a changing number of input signals has been used in these experiments. This set of experiments has been performed using only MLP classifier. The results obtained for the testing data representing all classes are shown in form of average accuracy in Table I.

TABLE I. Average accuracy of class recognition at different lengths of input vector.

Number of preceding months	MLP structure	Accuracy
6	160-10-3	99.1%
5	134-10-3	99.1%
4	108-10-3	98.2%
3	82-10-3	96.4%
2	56-10-3	95.9%
1	30-10-3	93.3%

The results show the significant role of the preceding months taking part in preparation of input data. The historical data representing a minimum of 5 preceding months seem to be enough, as the accuracy in testing data has achieved the maximum value.

Given the results presented in Table I, we have used 5 preceding months in input data preparation (134 input attributes) in further experiments. These data have been used in the final application of MLP and SVM classification systems.

Final experiments have been performed in 10-fold cross validation mode. The data have been split into ten parts: 9 parts have been used in learning and the remaining part only in testing. All learning and testing experiments have been repeated 10 times, exchanging the testing part of data each time. In this way, all 1500 samples of data have been used in the testing phase. The results of the experiments are presented in the form of a confusion matrix, on the basis of which the accuracy and sensitivity of a particular class recognition have been estimated. The confusion matrices corresponding to the application of MLP and SVM classifiers are presented in Table II. The rows represent target classes and columns represent the actual results of classification [16].

The results show that both classification systems demonstrate good quality in recognition of all classes. The best results have been achieved in recognition of active clients (100% accuracy in both systems). The level of accuracy of recognition of other classes (above 98%) is also fully acceptable in practice. In the case of recognition of churned clients, the MLP classifier was slightly worse.

TABLE II. Confusion matrices representing the performance of the system at application of MLP and SVM classifiers. The rows represent target classes (active, non-active, churned) and columns represent the actual results of classification

MLP

	Active	Non-active	Churned
Active	500	0	0
Non-active	0	499	1
Churned	3	6	491

SVM

	Active	Non-active	Churned
Active	500	0	0
Non-active	0	498	2
Churned	3	1	496

The statistical results in the form of average accuracy, sensitivity and precision in particular class recognition for both networks are presented in Table III. Sensitivity in class recognition, also called true positive rate, is defined as the ratio of the number of well recognized samples of the particular class to the number of all samples representing this class. On the other side, precision, also called positive predictive value, is the fraction of relevant instances of the particular class among the instances pointed by the classifier as the relevant class [16].

TABLE III. COMPARATIVE RESULTS IN THE FORM OF ACCURACY, SENSITIVITY AND PRECISION OF PARTICULAR CLASS RECOGNITION FOR SVM AND MLP NETWORKS (AVERAGE OVER 10 RUNS).

	MLP [%]	SVM [%]
Average accuracy [%]	99.3±0.797	99.6±0.786
Sensitivity (Churned) [%]	98.2	99.2
Sensitivity (Active) [%]	100	100
Sensitivity (Non-active) [%]	99.8	98.4
Precision (Churned) [%]	99.8	99.6
Precision (Active) [%]	99.6	99.6
Precision (Non-active) [%]	98.8	99.9

The detailed numerical results confirm the very high quality of the measures in recognition of all classes of clients, irrespective of the applied type of classifier. The important point in achieving these results is the correct choice of diagnostic features, delivered to the inputs of classifiers. Thanks to them, the system is well suited to such a task.

To check the influence of applied classification system the additional experiments have been done by using other classifiers. The application of random forest of decision tree has resulted in average accuracy equal 98.7% and logistic regression 97.5%. Such good results are due to very good choice of diagnostic features, defined for this particular problem. Moreover, the significant role was also balance of all classes.

The additional experiments have been also performed for unbalanced set of data. This time the number of records representing active class was 6726, non-active 1219 and churned 2279. Due to very high number of cases the destinations of each record were done half-automatically (without personal intervention of expert). The records have been associated with active state for companies with nonzero bank account and vivid operations within the last months. Non-active state was assigned to companies of small total volume of product balance within the last months. Status churned was automatically associated with clients, who closed physically the account.

The same structure of classification system arranged in an ensemble has been applied. This time the average accuracy for such set was 86.7%. The sensitivity of recognizing the active and non-active cases was quite high: 95.9% and 91.3%, respectively. The lowest sensitivity achieved the class of churned, which was misclassified with non-active. The sensitivity of churned was only 73.2%. In our opinion the most important reason of decreasing this accuracy was non-perfect destinations associated automatically by the system, without confirmation of the bank expert.

VI. CONCLUSIONS

This study shows that the data mining techniques based on proper definition of input attributes and application of artificial neural networks provides a good tool for supporting the prediction of customer behaviour in corporate banking. The results of such predictions will give the bank a chance to undertake some actions to prevent the churning decision of customers.

The developed computer model of customer behaviour applies the selected numerical data related to the previous activity of customer. On the basis of these data, the set of unique diagnostic features that can serve as the input attributes to the classification system, responsible for the prediction of customer decision, have been elaborated. These features represent the dynamic changes of the selected features within the considered past time period.

Our results show that the proposed dynamic behavioural model related to a few preceding months could predict the customer behaviour and possible churning decisions in a precise way, equal or even better than the traditional way provided by bank experts. An additional study on the relative importance of the past behaviour of customer has shown that 5 months' worth of observations are enough to predict the future decision of the customer of corporate banking with sufficient accuracy.

The important advantage of the method is its absolute repeatability of results, in contrast with the human expert results, which are considerably dependent on the particular choice of the expert and his actual mental and physical condition. Moreover, our fully computerized system provides a very quick operation, especially in comparison to the time required by an expert to study a huge amount of past data related to many clients of the bank.

REFERENCES

[1] W. Buckinx and D. Van den Poel, "Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting", European Journal of Operational Research, vol. 164, pp. 252-268, 2005.

[2] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees", European Journal of Operational Research, vol. 269, no 2, pp. 760-772, 2018.

[3] Z. Y. Chen, Z. P. Fan, and M. Sun, "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data", European Journal of Operational Research, vol. 223, no 2, pp. 461-472, 2012.

[4] S. K. Depren, "Churn propensity model for customers who made a complaint in retail banking", European Journal of Business and Social Sciences, vol. 6, no 11, pp. 48 – 61, 2018.

[5] V. P. Eswaramurthy, and Induja, "A study on customer retention using predictive data mining techniques", International Journal of Computer & Organization Trends, vol. 4, no 5, pp. 6-10, 2014.

[6] N. Glady, B. Baesens and C. Croux, "Modeling churn using customer lifetime value", European Journal of Operational Research, vol. 197, pp. 402-411, 2009.

[7] B. He, Y. S. Qian and W. X. Zhao, "Prediction of customer attrition of commercial banks based on SVM model", Procedia Computer Science, vol. 31, pp. 423 – 430, 2014.

[8] B. Q. Huang, T. M. Kechadi, B. Buckley, G. Kiernan, E. Keogh and T. Rashid, "A new feature set with new window techniques for customer churn prediction in land-line telecommunications". Expert Systems with Applications, vol. 37, pp. 3657-3665, 2010.

[9] R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques", Applied Soft Computing, vol. 24, pp. 994–1012, 2014.

[10] E. Kaya, X. Dong, Y. Suhara, S. Balcisoy, B. Bozkaya, and A. Pentland, "Behavioral attributes and financial churn prediction". EPJ Data Science, vol. 7, no 1, 2018, https://doi.org/10.1140/epjds/s13688-018-0165-Keramati.

[11] D. Popović and B. D. Bašić, "Churn prediction model in retail banking using fuzzy c-means algorithm", Informatica, vol. 33, pp. 243-247, 2009.

[12] U. D. Prasad and S. Madhavi, "Prediction of churn behavior of bank customers using data mining tools", Business Intelligence Journal, vol. 5, no 1, pp. 96-101, 2012.

[13] R. Rajamohamed and J. P. R. Manokaran, "Improved credit card churn prediction based on rough clustering and supervised learning techniques", Cluster Computing, 2017, DOI:10.1007/s10586-017-0933-1.
[14] Schölkopf B., Smola A., "Learning with kernels", MIT Press, Cambridge MA, 2002.
[15] J. Sun, K. Y. He and H. Li, "SFFS-PC-NN optimized by genetic algorithm for dynamic prediction of financial distress with longitudinal data streams". Knowledge-Based Systems, vol. 24, pp. 1013-1023, 2011.
[16] Tan P.N., Steinbach M., Kumar V., "Introduction to data mining", Pearson Education Inc., Boston, 2006.
[17] C. F. Tsai and Y. H. Lu, "Customer churn prediction by hybrid neural networks", Expert Systems with Applications, vol. 36, pp. 12547–12553, 2009.
[18] D. Van den Poel and B. Larivière, "Customer attrition analysis for financial services using proportional hazard models". European Journal of Operational Research, vol. 157, pp. 196-217, 2004.
[19] W. Verbeke, K. Dejaeger, D. Martens, J. Hur and B. Baesens, B.,"New insights into churn prediction in the telecommunication sector: A profit driven data mining approach". European Journal of Operational Research, vol. 218, pp. 211-229, 2011.
[20] A. Wangperawong, C. Brun, O. Laudy and R. Pavasuthipaisit, "Churn analysis using deep convolutional neural networks and autoencoders", 2016, arXiv preprint arXiv:1604.05377
[21] Y. Xie, X. Li, E. W. T. Ngai and W. Ying, "Customer churn prediction using improved balanced random forests", Expert Systems with Applications, vol. 36, no 3, pp. 5445–5449, 2009.