

Data Mining Applications in a Financial Sector-Case Study of Czech Republic

Mohelska Hana and Vokalova Andrea
Faculty of Informatics and Management, University of Hradec Kralove,
Hradec Kralove, Czech Republic

Abstract: Data mining is a process of sifting through a very large amounts of data for useful information. Data mining uses artificial intelligence techniques, neural networks and advanced statistical tools to reveal trends, patterns and relationships which might otherwise have remained undetected. Currently, this information is used increasingly, despite the significant difference in the way that organizations handle such information. This study aims to describe data mining in the financial sector by using two typical banks which together hold more than 7.8 million clients in the Czech financial sector. A common feature of these companies is their size and total number of clients. The study evaluates the role and importance of data mining in these large banking institutions. The first research question of the presented study is dealing with a utilization of DM and its influence on profit and on the total number of clients. The second research question is evaluating the level of client's segmentation and its advantages in both banks. Generally, the Czech financial sector uses DM very sophisticated in order to not only evaluate its actual state (thank to reporting, etc.) but also in order to help the predictions of future needs and state.

Key words: Data-mining in financial sector, period of business understanding, period of data understanding, CRM, segmentation, State

INTRODUCTION

Data Mining (abbr. DM) is an evolving and growing area of research and development both in academia as well as in industry. It involves interdisciplinary research and development encompassing diverse domains. In our view, this area is far from being saturated with newer techniques and directions being proposed in the literature every day. In this age of multimedia data exploration, data mining should no longer be restricted to the data mining of knowledge from large volumes of high-dimensional datasets in traditional databases only. Researchers need to pay attention to the data mining of different data types, including numeric and alphanumeric formats, text, images, video, voice, speech, graphics and also their mixed representations. Efficient management and new approaches to such high-dimensional very large databases also influence the performance of data mining systems. Knowledge data compression and understanding can play a significant role. It is also important that special multimedia data compression techniques are explored, especially, suitable for data mining applications (Mitra and Acharya, 2003). Business analysts working for Large Enterprises (abbr. LEs) in the Czech financial sector are using data mining for evolving new trends, patterns and indicators that can improve the increase in sales, IT and other operations.

The characteristics of the Czech financial sector: Basic characteristics of financial and mainly banking sector in the legal environment of the Czech Republic are:

- The financial sector is clearly dominated by commercial banks
- In terms of the ownership structure the Czech financial sector is highly internationalized
- The overwhelming majority of all banks is here represented by large banking institutions
- Despite of the high number of large banks, still remains a fairly significant degree of mutual competitive behavior between them, although, it is possible to perceive the distinctions in different segments of the banking
- Medium-sized banks constantly enhance their activity and represent one of the driving forces focused on a continuous stream of market innovations and also because the large banking houses are real competition in some selected segments of the banking market
- Banks are leaders in almost all financial groups operating in the Czech Republic, so, it can be simplified and used in the term financial market, where banks play a major role led by financial conglomerates (Berka, 2003)

The financial sector in the Czech Republic is supervised and regulated by the Czech National Bank (CNB) (Ceska Sporitelna, 2016; Czech National Bank, 2017) which since April 2006 is an institution in which they are concentrated surveillance activities over the whole financial market and all its subsegments, i.e., excluding banks themselves also (Everitt *et al.*, 2011):

- Credit unions
- Insurance companies
- Pension funds
- Investment companies and unit trusts
- Leasing companies, etc.

In recent years, banks pay attention to the development of retail banking which includes household's loans (not only consumer credits) but also mortgages. Retail loans create quite a dynamic segment of the banking market (Herawan and Deris, 2011).

The definition of data mining: The concept of Data Mining (abbr. DM) comes from English and translates as mining in a large amount of data. DM itself uses a number of techniques, procedures and algorithms used to discover meaningful correlations, patterns and trends during the process of browsing large amounts of data frequently stored in databases or in data warehouses. Basic and commonly used definition of DM is: "Data mining is also called Knowledge Discovery in Databases (KDD). It is commonly defined as a process of discovering useful patterns or knowledge from data sources, e.g., databases, texts, images, the web, etc. The patterns must be valid, potentially useful and understandable. DM is a multi-disciplinary field involving machine learning, statistics, databases, artificial intelligence, information retrieval and visualization. There are many DM tasks. Some of the common ones are supervised learning (or classification), unsupervised learning (or clustering), association rule mining and sequential pattern mining" (Herawan and Deris, 2011).

DM is applied in many sectors, for example, to evaluate the return of campaigns, analyses of customer's database, search for broader context and possible correlations in an effort to find the data from all the things that can help the management to lead the company, project or employees (Herawan and Deris, 2011). Thanks to DM system, the data can also be easily processed through the recognition technologies as well as graphical and statistical techniques. Thus, allowing the data presented in a comprehensible form useable for public or for the contractor. However, DM can be applied to small databases data (Jayasree and Balan, 2013). An

important feature of DM is a fact that the analyses are derived from the content of obtained data not predetermined by the user. These are mainly about deducing predictive information, not merely descriptive ones. Compared with other statistical tools, DM is advantageous for its focus on the user. DM statistical tasks are performed automatically according to specific algorithms and users do not need to have special knowledge of statistics. A specialized user creates reports for the contractor in other software (Slik, 2016; Czech National Bank, 2017).

Basic sources of data for data mining: There are defined three basic data sources: databases, statistics and machine learning.

Databases: This group includes all the systems in which is data collected and which enable developing specific group of data. Very often it is a special relational database (abbr. DBMS) in which data is organized for the acceleration of the analytical operations (e.g., do not meet 3NF) that provide functionality necessary for an analysis, such as OLAP (Online Analytical Processing). These systems are called Data Ware Houses (DWH). Operations and questions in these systems are usually done using SQL query language or its other modifications. Furthermore, this group also includes EIS systems (Executive Information System).

Statistics: Statistics presents tools for mathematical data processing such as Pivot table, regressive analysis, distribution and correlative analysis (Weikum *et al.*, 2009).

Machine learning: Machine learning is a modern, constantly evolving technology used for knowledge discovery and data acquisition. It uses different algorithms for classification of existing data in order to detect trends or predictions of new data. This category also includes the so-called genetic algorithms based on biological evolutionary process or on very reassuring neural networks. Examples are management theory and artificial intelligence (Weikum *et al.*, 2009).

A brief history of knowledge discovery in databases: Essentially, the first impulse which preceded the formation of data mining can be dated to the 30's and 40's of the 20th century. At this time SAS and SPSS were dealing with mathematical analysis such as standard deviations, analysis of variance, cluster analysis and confidence intervals. The first methods aimed to explore databases and search for hidden connections among this data began to emerge about 50 years ago. In the early 60's there was

an important development of computer technology which lead to the idea of getting a useful information from stored data. But mostly there was an isolated and sporadic research administered at academic level (Berka, 2003). In the 70's and 80's was the development supported by increasing the computational power and speed of computers. Thank for this, there was a significant expansion of using statistical methods, database applications and artificial intelligence. Resources from this era described the process of data mining as a process of data extraction, uncovering information and knowledge acquisition. These terms were used in a rather derogatory meaning because it was not possible to ensure the reliability of the results. Furthermore, it was more about finding correlations in large data sets. The term of the real process of Knowledge Discovery from Databases (KDD) was used in specialized literature in the early 90's of the 20th century. Firstly, this term was introduced in the United States where in 1989 the term was established during the international conference on Artificial Intelligence (IJCAI'89). KDD is defined as a "non trivial process of discovering implicit, unknown and potentially useful knowledge in data" (Yap *et al.*, 2011). In the same year, the first KDD workshop emphasized the need to get useful knowledge from databases. Therefore, until the 90's, after the discovery of new statistical methods, we can clearly speak about data mining as a separate and equal field of applied science.

MATERIALS AND METHODS

The study is based on methodology called CRISP-DM which aims to find solution and important knowledge discovery from large databases (such as Customer Relationship Management). There are specified typical tasks of data mining in the Czech financial sector, which are examined in the study. The first research question of the presented study is dealing with a utilization of DM and its influence on profit and the total number of clients. The rise of net profit and clients higher than 10% confirm the statement that DM can affect these indicators. This is followed by the original case study of the two largest Czech banks-Ceska Sporitelna, PLC (abbr. CS) and CeskoSlovenska Obchodni Banka, PLC (CSOB) (CSOB Group, 2012). Together, these banks share more than 7.8 million of clients (more than 75% of clients) on the Czech financial market. Case study is dealing with large amount of various data associated with these clients which are stored in bank's CRMs. KDD helps to innovate bank's sales strategy and also provides new approaches or trends for better client's segmentation. The second research question is evaluating the level of client's

segmentation in both banks. The last part of the study focuses on the advantages and disadvantages that provides applying DM and KDD in data mining and the bank needs for experienced DM analysts recruiting.

Typical tasks of dm in the financial sector: The critical moment of DM is the implementation and integration of selected solution into the company's operational system. DM derived models and procedures must be implemented. Implementation can be supported by using DM tool, if it is possible. Typical tasks of data mining in the commercial sector are:

- Fraud detection in the insurance or banking industry (fraud)
- Identification of clients who want to leave the bank to competitors (churn)
- Client's segmentation for marketing purposes (cross/up-sell)
- Analysis of shopping basket in a shop and others

The common feature of these tasks is a fact that they are either once or repeatedly making an evaluation of a subject-loan, client, mortgage, credit score, insurance report, etc. The subject is assigned by the score of probability/this probability says the percentage that means how possible is the examined situation to occur (such as leaving to competition, bankruptcy, outstanding debts, fraud insurance event, positive feedback on a marketing campaign), etc. An important factor is the time, when this knowledge or this score must be available and how actual it must be at the dedicated moment. Therefore, it is necessary to distinguish these tasks in following groups (Berka, 2003):

- On-line
- Off-line

As for on-line, real-time tasks it is necessary to know the score immediately. When the client is now at the counter or on the phone, banker must have maximum information for a correct advise or decision, whether to approve a loan, credit card, mortgage or offer a different product or try to increase client's loyalty by offering another benefit. Off-line tasks, calculate score mostly in doses for a large number of cases. These calculations are mainly processed at a time when the database servers are not used or busy, typically overnight and in required frequency (such as daily, weekly, monthly).

Segmentation is the process of partitioning markets into groups of potential customers with similar needs and or characteristics who are likely to exhibit similar purchase

behavior. It has emerged as a key marketing planning tool and the foundation for effective strategy formulation in American and international companies. The objective of segmentation research is to analyze markets, find niche opportunities and capitalize on a superior competitive selection. This can be accomplished by selecting one or more groups of users as targets for marketing activity and developing unique marketing programs to reach these prime prospects (market segments) (Weinstein, 2013). Basic two segments defined in all banking sectors are private clients (people) and corporate clients (entrepreneurs and companies). Thank to DM, some banks have successfully defined other market segments which are described in following case study.

RESULTS AND DISCUSSION

Case study of DM utilization in selected banks: The part of presented case study is aimed at DM applications, used methodology or methods in field of segmentation in the largest bank group operating on the Czech financial market. Ceska Sporitelna has a long history in the Czech Republic and the main characteristics of CS are: “CS is a modern bank with a focus on retail clients, small and medium-sized enterprises, municipalities and cities and we play an important role in financing large corporations and providing financial services. With more than 5 million clients, we are the Czech Republic’s largest bank. We boast the longest tradition among the banks on the Czech market in 2015 we mark the 190th anniversary of the company’s founding. In Fincentrum’s Bank of the year competition, we received the Most Trustworthy Bank of the Year award for the 11th year running. The prestigious magazine *The Banker* selected CS as the Bank of the Year 2014 which is the eighth title won by the company. We provide our clients with the broadest range of banking services in the Czech Republic through a branch network, ATMs and payment terminals (The Czech Banking Association, 2016).

Department of DM in CS is a part of its analytical and reporting division. Many analysts seek for new trends and pattern of the client’s behavior and needs. The data sets for evaluation are transferred from CS CRM system called Siebel Klient. Since, 2010 the CS databases have grown and the data stored in them as well. There was also a significant change in retail banking sales strategy that was caused by small banks (such as Fio Bank, Mbank, etc.). The management of CS was forced to come up with a plan or project that was crucial for a better understanding of clients needs and wishes and also prevents their leaving. The new project should find these client’s characteristics in their datasets or any other

information stored in CRM. This was a goal for CS data mining analysts, who should provide new patterns from deep data mining by using approaches as introduced CRISP-DM. Thank to their activity and analytical skills was created a new and quite a unique project called “Color bank” or sometimes called color segmentation. The color segmentation project was created in order to provide better understanding of client’s behavior, understand their preferences and meet their needs and wishes.

This project is primarily focused on accomplishing the business strategy and objectives. The secondary goal was to be or become a relevant and trustworthy partner for clients. These goals can be enabled only by an active approach to clients an effort to fully understand them, respect their requirements and offer them an appropriate solution. This does not mean of course that the bank has to change its nature and pretend to be something else. The goal is to understand what clients really want and why. The point is to realize the broader context of different lifestyles, understand the uniqueness of each individual client. The project color segmentation was working on five defined goals which are:

- More detailed client’s segmentation within the mass market segment
- More accurate identification of a client’s financial needs in selected groups
- Determining appropriate service models for individual segments and their possible crossings
- Determination of the basic elements of successful communication strategy for each segment
- Getting a basic understanding of how to target better product offers

According to these goals and with the help of DM department was established a new segmentation model used in CS. This model is evolving the two mentioned segments (e.g., private and corporate clients) and divides potential clients into 4 color groups. The first color group is beige and this segment is developed for majority of clients, who seek for simple financial solutions and their values are friendliness, simplicity and trust. The second segment is focused on wealthier clients, who earn more than 30,000 CZK monthly or have 0.5 million CZK on their account. This segment is blue and target on convenient financial solutions for ensured clients, whose values are professionalism, responsibility and partnership (Fig. 1).

Third segment is focused on the wealthiest clients and its color is black. Last (grey and the 4th segment) is aimed on entrepreneurs and small or medium-sized enterprises. This project was developed in 2011. Year later

Blur clients	
Who? for whom?	30,000+monthly or quota 0.5+million
What?	Convenient financial solutions for ensured clients
Values?	Professionalism, responsibility, partnership

Fig. 1: Basic characteristics of 2nd segment-BLUE clients

in 2012 was this concept implemented in the CS marketing campaigns and brought more than 94,000 of new clients (The Czech Banking Association, 2016). The total number of clients was 5.202 million in 2011. In 2012 the number of clients rised to 5.292 million. Net profit of CS was 13.6 billion CZK in 2011. After implementation of the project color bank the net profit increased to 16.6 billion CZK in 2012 (more than 3 billion increase in 1 year). In case of CS we can accept the 1st research question and confirm that DM and its successful application (due project color bank), can positively affect the total number of clients as well as net profit. In CS annual report from year 2012 is also mentioned that DM can support bankers because after implementation the project color bank segmentation, it is easier to understand clients, their needs and there is higher probability to offer them a suitable product or service. Second research question is dealing with the level of client's segmentation in selected bank. CS has its own DM and analytical department where employees are regularly trained and work on new projects. The biggest disadvantage is a lack of DM specialists in CS. That is why the 90% of them are CS employees with analytical skills and banking experience who visit the DM department for internship and some stayed permanently. The 2nd research question agreed with the CS annual report and confirmed that the level of client's segmentation in CS is very sophisticated and constantly developed.

The second largest bank operating in the Czech financial sector is CSOB. "Ceskoslovenska Obchodni Banka, PLC (CSOB) is a Universal Bank in the Czech Republic. CSOB was established by the state in 1964 as a bank for providing services in the Field of Foreign trade financing and convertible currency operations in the then Czechoslovak market. It was privatized in 1999-KBC Bank which is a 100% subsidiary of an integrated bank-insurance group KBC Group, became the majority shareholder (since, 2007 the sole shareholder). In June 2000, CSOB took over the Bank Investicni and Postovni Banka (IPB). Until the end of 2007, CSOB was active in the Czech and Slovak markets, the Slovak branch of CSOB

was transformed into an independent legal entity on 1st January 2008. Effective as of 1st January 2013, KBC Group has organized its core markets activities into three business units-Belgium, Czech Republic (includes all KBC business activities in the Czech Republic) and international market. CSOB has its own department focused on analytical and statistical methods which integrated DM analysts. As for segmentation, there is still used quite outmoded approach and defined concept. Basic two segments defined in all banking sectors are private clients (people) and corporate clients (entrepreneurs and companies). The segment is then divided into children, students, adults and seniors. The segment of corporate clients is divided into small, medium-sized and large enterprises. This structure of segments is quite outdated and there are no new projects, in order to make it easier and better (not only for potential clients) but also for their employees. The net profit of CSOB was 11.172 billion CZK in 2011 and in 2012 increased to 15.282 billion CZK. These can be affected by the increasing number of mortgages in the Czech Republic but CSOB could gain higher profit if there was a better system of client segmentation. The total number of clients is falling which is a great opportunity for competitors (such as CS and others). In 2011 the total number of clients was 3.096 million. On the other hand in 2012 it was 3.054 million clients. The difference in one year is 42,000 clients which is quite enormous number and the management of the CSOB should work with the DM analyst in order to improve this state. The 2nd research question was focused on the level of segmentation in selected banks. CSOB has poor segmentation model (as well as its level) and should improve it significantly in following years in order to compete with CS. As for the level of client's segmentation in CSOB we can declare that compared to CS, it is quite outdated and rigid. CSOB should use the large amount of client's data in their CRM and with the help of DM develop a new segmentation model which will help not only potential clients but also bankers.

CONCLUSION

Data mining provides tools and techniques which are used to discover important and meaningful information from the large datasets (such as CRMs or databases). This process enables banks to make better decisions and also better understanding of client's needs and wishes. These understandings in the financial sector can help banks with a better segmentation of clients, also with reduction of costs and using the data from CRMs. Banks and Financial organizations started allocating funds and

time for implementing data mining tools in the area of decision making by realizing the necessity of data mining in their system. Quite important is the question of recruiting a specialized data mining analyst, who are very hard to find and employ. Often must recruiters search for their own employees with analytical skills or experience and give them a DM training in order to make them future DM analysts.

The 1st research question shows that through DM can banks increase their profit, number of clients and their loyalty (approved by CS) and also employee satisfaction. The 2nd research question is evaluating the level of client's segmentation in both banks. CS is much better in this comparison because there are implemented and developed many new projects (such as project color bank) and approaches associated with DM applications. CSOB has a quite insufficient segmentation model and must work on it in future.

ACKNOWLEDGEMENTS

This study was supported by the research project SPEV (Determinants that have an impact on job satisfaction), University of Hradec Kralove, Faculty of Informatics and Management, 2016.

REFERENCES

- Berka, P., 2003. Extraction Knowledge of Databases. Academia, Praha, Slovakia.
- CSOB Group, 2012. Czechoslovakia of commercial banks. CSOB Group, Prague, Czech Republic. <https://www.csob.cz/portal/about-csob/about-csob-and-kbc/about-the-csob-group>.
- Ceska Sporitelna, 2016. Profile of Ceska Sporitelna the modern bank with the longest tradition. Czech Republic. http://www.csas.cz/banka/content/inet/internet/en/STANDARD_CONTENT_OT01_008647.XML.
- Czech National Bank, 2017. The central bank of the Czech Republic and the supervisor of the Czech financial market. Czech National Bank, Czech Republic. <https://www.cnb.cz/en/>.
- Everitt, B.S., L. Sabine, L. Morven and D. Stahl, 2011. Cluster Analysis. 5th Edn., Wiley, Chichester, England, UK.,.
- Herawan, T. and M.M. Deris, 2011. A soft set approach for association rules mining. Knowl. Based Syst., 24: 186-195.
- Jayasree, V. and R.V.S. Balan, 2013. A review on data mining in banking sector. Am. J. Appl. Sci., 10: 1160-1165.
- Mitra, S. and T. Acharya, 2003. Data Mining: Multimedia, Soft Computing and Bioinformatics. John Wiley and Sons, New Jersey, ISBN: 9780471460541, Pages: 401.
- Slik, L., 2016. Data mining banks. Magazine of IT System. <http://www.systemonline.cz/business-intelligence/data-mining-v-bankach.htm>.
- The Czech Banking Association, 2016. The banking sector. The Czech Banking Association, Czech Republic. <https://www.czech-ba.cz/cs/bankovni-sektor>.
- Weikum, G., G. Kasneci, M. Ramanath and F. Suchanek, 2009. Database and information-retrieval methods for knowledge discovery. Commun. ACM., 52: 56-64.
- Weinstein, A., 2013. Handbook of Market Segmentation: Strategic Targeting for Business and Technology Firms. 3rd Edn., Routledge, New York, USA., ISBN:0-7890-2156-0, Pages: 223.
- Yap, B.W., S.H. Ong and N.H.M. Husain, 2011. Using data mining to improve assessment of credit worthiness via credit scoring models. Expert Syst. Appl., 38: 13274-13283.