Data Mining Applications for Banks' Business Intelligence Maturity

Mina Ranjbarfard, Shahideh Ahmadi
Alzahra University
Iran
{m.ranjbarfard@alzahra.ac.ir} {shahideh.ahmadi@gmail.com}

JDIM
Journal of Digital Information Management

ABSTRACT: *There are many researches that applied data mining in banking. However, one may be countered the lack of suitable data as a serious obstacle to employ data mining techniques for the banks. This paper examines previous data mining researches carried out for banking, integrates them, extracts all served entities and attributes which are needed for analytical purposes, categorizes them and ultimately presents a suitable data model for analytical purpose. After analysis of a wide range of data mining application in banking, 28 entities with 423 attributes were identified in conclusion and the final proposed entity-relationship diagram was created. Also a checklist was provided based on the model for auditing data gap in banks and applied in a real case. The results of this paper can be regarded as a supportive tool for increasing banks' business intelligence maturity from the data perspective and enable managers in requirement analysis of information systems.*

Subject Categories and Descriptors:
[H.2.8 Database Applications]: Data mining; **[J.1 Administrative Data Processing]:** Financial **K.4.4 [J.1Electronic Commerce]** Electronic data Interchange (EDI)

General Terms: Data Mining, Data in Banking, Electronic Commerce

Keywords: Data Mining, Banking, Data Model, Entity-relationship Diagram

Received: 19 March 2020, Revised 4 July 2020, Accepted 19 July 2020

Review Metrics: Review Scale: 0/6, Review Score: 4.9, Inter-reviewer consistency: 78.2%

DOI: 10.6025/jdim/2020/18/5-6/163-172

1. Introduction

Banking is a fertile industry for new information systems and technologies [1]. among them are business intelligence and decision support systems. Business intelligence as an important concept of machine learning and big data [2] is a term that includes architectures, tools, databases and methodologies for analyzing data in order to support decision making for business executives [3]. Banking domains such as branch efficiency, e-banking, customer segmentation and retention, and etc. provide an extraordinary context for broad application of business intelligence concepts and its methods such as data mining, data warehouses and decision support systems [4], [5], [6]. Data mining methods of business intelligence have been applied for improving banking operations such as fraud detection [7], [8], [9], credit assessment [10], [11], [12], customer churn prediction, and etc. [13]. Nowadays, banks are aware about the value of their customer related data. They need to transform their daily transactions data for complex analysis including risk management, customer relationships, profit and production channels, market valuation, operational efficiency, etc. [14]. Failure in these areas results in undesirable output such as loss of customers, property damage and loss of credit and heavy fines [6]. However, one may be counter when analyzing data that the required data attributes are not consolidated and organized in order to perform the required analysis. In order to create knowledge in these areas, the relevant data gathered by the information systems is necessary and the lack of suitable data is a serious obstacle to employ data mining techniques for the banks. However, so far, no research in the field of banking has addressed this issue that what information items are needed to take advantage of data mining applications in the banks? And basically what is the suitable data model in banks to maximize usage of analytical applications? And how can one assess richness of existing data in a bank with the aim of applying data mining techniques? Therefore, it was sensed to examine the various research about data mining application in banking and extract their served data attributes and their relationship in the form of a data model so that banks find what are necessary data to provide in their information systems development future plans in order to keep moving with the latest progress in intelligent banking. This research seeks to identify various applications of datamining in banking and relevant data gaps in banking information systems to provide a comprehensive data model for maximizing the usage of data mining and business analysis applications. This model helps the bank‘s IT managers to consider the analytical requirements in long-term planning of information systems development so that required infrastructure for data mining applications and analysis capabilities will be provided in their bank. The final presented data model applied a checklist in a bank for auditing data gap as a sample for the application of the data model. Also, the results of this paper can be regarded as a supportive tool for increasing bank‘s business intelligence maturity from the data perspective.

2. Theoretical Literature and Research Background Datamining in Banking

Data mining is the process of analyzing and summarizing data from a variety of perspectives. This field involves new theories and methods for processing large volumes of data [15]. Data mining is a repetitive process that combines business knowledge, methods, machine learning tools and large amounts of correct and relevant information to uncover hidden unseen perspective of the organization’s data. This information can correct the existing processes and show trends to adopt customers and employee’s policies [16].

Banking industry has noticed the importance of information about customers that is undoubtedly among a large volume of customer information, demographic information, transactional data, credit card usage patterns and so on. Banks can offer customized products and services to customers by using suitable data mining tools [5]. Banking systems collect a large amount of information every day which includes customer information, transaction details, risk profiles, credit card details, collateral details, compliance and anti-money laundering information and Swift messages. Thousands decisions are decided on a daily basis. These decisions include issues of credit, fraud, investment, money laundering and illegal financial. A manager needs to get different reports and be able to make critical decisions by various banking system tools [6]. Using data mining to analyze patterns and trends, bank managers can predict accurately how do customers react to new rates or which customers are likely to accept new products.

There have been a several application of data mining in banking. For example, used Logistic regression, Classification and Regression Tree (CART) and Cascade Correlation Neural Network (CCNN) to build knowledge-based scoring models [17; 18]. displayed that how big data analytics is being successfully used in banking sector in some aspects including spending pattern of customers, channel usages, customer segmentation and profiling, product cross selling, sentiment and feedback analysis, and security and fraud management. Credit risk is the main application trend in banking. There is also a relevant interest in bankruptcy and fraud prediction. After that, customer retention seems to be associated, although weakly, with targeting, justifying bank offers to reduce churn [4].

Credit risk assessment for secured loans is an important operation [19]. with the aim of identifying necessary factors to assess credit risk, generated a new decision tree model based on C 5.0 methodology in order to reduce the of number of nonperforming loans [20]. developed some Monte Carlo experiments using known techniques and algorithms and implemented a linear mixed model (LMM) as a new contribution to calculate the credit risk of financial companies [21].

Studied determinants of deposit pricing by employing various methods including multivariate adaptive regression splines, support vector regression, artificial neural networks, classification and regression trees and random forest results highlighted the importance of customer- and account-specific characteristics in the determination of deposit rates. Also it was indicated that depositors with a multi-faceted and long-term relationship with the same bank seem to benefit from higher deposit rates as a reward for being a core depositor [22]. proposed a model for predicting liquidity risk using Artificial Neural Networks and Bayesian Networks. Two-phase method. Their model predicts liquidity risk by identifying its triggering factors [5]. examined support vector machines and random forests with logic regression to detect fraud using the real data of international credit card payment transactions. Random forests showed better performance than other techniques, though logic regression and support vector machines also worked well [23]. on the study of the Taiwan Bank were used Bayesian classification and association rule to identify the signs of fraudulent accounts and the patterns of fraudulent transactions. Detection rules were developed based on the identified signs and applied to the design of a fraudulent account detection system. Using clustering techniques on money transferring data in Vietnamese bank [24], proposed some approaches on money laundering detection techniques. They presented a system for detecting money laundering using CLOPE algorithm [25].

On the study of the Bank of England, proposed a model based on fuzzy logic and data mining algorithms to characterize the e-banking phishing website factors. The techniques evaluated by classifying the various phishing types and defining six criteria for attack the e-banking phishing website in a layered structure. In a Spanish Bank case study [26], reviewed various methods and techniques to determine the important variables required for financial institutions to predict the likely levels of trust among e-banking users. To do so, the most recent advances in machine learning and soft-computing were used, including a new selection operator for multi objective genetic algorithms. Their new proposed methodology, obtained the best results in terms of optimization and highest punctuation by the experts [13]. In a case study of a European private bank, proposed a dynamic churn prediction framework for generating training data from customer records and leveraged it for predicting customer churn within multiple horizons using standard classifiers [27]. in a case study of Portage Bank, proposed a data mining response model supported by random forests to support the definition of target customers. The performance of an underdamping method (the Easy Ensemble algorithm) was compared with that of an oversampling method (the Synthetic Minority Oversampling technique). The importance of the attribute features included in the response model was also explored. Random forests that were supported by an underdamping algorithm, presented very high prediction performance compared to the other techniques [28]. on the study of the Turkish bank, developed an applicable and detailed model for customer lifetime value (CLV). The results of the least square estimation (LSE) and artificial neural network (ANN) was compared in order to select the best performing forecasting tool to predict the potential CLV. Due to its higher performance; LSE based linear regression model was selected. The proposed model included not only profit and costoriented indicators, but also certain other indicators rarely were used in the literature [29]. looked at both the account data of the customers and their credit card transactions in the study of the Taiwan Bank. The aim was to discover interesting patterns in the data that could provide clues about what incentives a company could offer as better marketing strategies to its customers. A two-stage behavioral scoring model was presented with a cascade involving selforganizing map (SOM) and an Apriori association rule inducer [30]. applied a fuzzy data mining technique called Fuzzy Association Rule Mining II (FARM II) to help Hong-Kong County Bank for better serving and retaining customers through discovering customer‘s hidden patterns [31]. applied particle swarm optimization (PSO) to obtain suitable parameter settings for support vector machine (SVM) and decision tree (DT), and select a subset of beneficial features in order to predict bank performance. Experimental results showed that their proposed approaches improved the accuracy of classification significantly [32]. predicted bank failure in the U.S. banking sector using extreme gradient boosting (XGBoost) utilizing annual series of 30 financial ratios for 156 U.S.

National commercial banks [33]. applied Extreme Gradient Boosting on the 25 annual financial ratio series data to identify a set of key indicators that help predict and prevent bank failure in the Eurozone banking sector. In a study by the Bank of America [34], used four data mining methods including logistic regression, decision tree, neural network, and knearest neighbor in order to predict bank failure [15], in the case of the 102 branches of a large private bank, presented an intelligent integrated decision support system (DSS) by integration of data mining tools (RST, ANN, MLP, GA, and CVTT), DEA and KMeans for prediction and optimization of complex personnel efficiency [35]. proposed a data warehousing architecture for effective risk analysis in banking. Furthermore, they presented a hash based technique for data reconciliation. On the study of an Iranian bank [36], classified 18 branches of a certain bank in order to determine sufficiency of cash in bank‘s branches. Using hierarchical clustering and Bayesian hierarchical clustering in similar clusters, they estimated the amounts of entered and consumed branch cash through neural network.

3. Data Model and Entity

The data model is a method and tool for defining realworld information requirements to be understandable for the organization ‘s shareholders. In addition, data modeling enables database specialists to use the information requirements for implementing a computer database system in order to support the organization. Therefore, a data model is a critical tool for communication with users and it also provides a database system design for developers [37]. Entity type conceptually is like the concept of a class in object-oriented design. An entity type can represent a set of people, places, objects, events or concepts. Examples of entity type in an ordering system can include customer, order and tax [38]. An entity is represented by a set of attributes. Attributes are descriptive properties that each entity possesses, such as name, last name, address and telephone for customer entity [39]. The ER[1] model is one of the best known tools for database logic design. This diagram is also understandable by nonprofessionals, is easily visualized, their entities and their relationship are visual, thus is a natural way to display the user’s information needs. Database systems are typically modeled by using an entityrelationship diagram as a design of the stored real data [40].

4. Research Method

This research in terms of purpose is an applied research [41]. because it seeks for an entity relationship data model to identify the current data gap in banks and financial institutions. By using this model, it can be determined which entities and attributes are currently banks and financial institutions, and what used in are the gaps.

In terms of the implementation process, the present research is qualitative [42]. In this research, content analysis method has been used. The content analysis method

[1] Entity Relationship Model

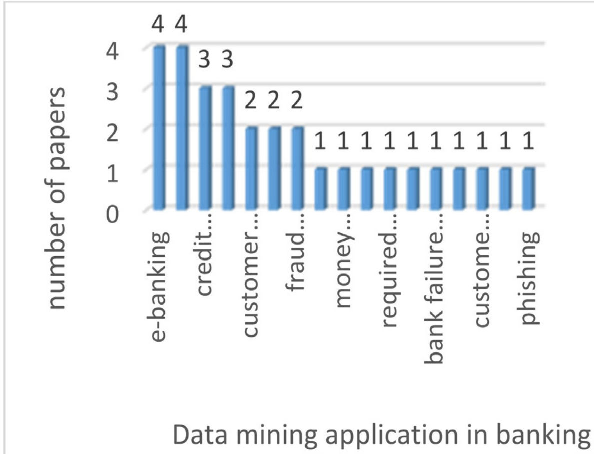

Figure 1. Frequency diagram of data mining applications

Organizational Position	Sex	Age	Experience	Education	Raw
Expert	Male	46	19	M.Sc.	1
Department head	Male	43	18	M.Sc.	2
Expert	Female	36	10	B.Sc.	3
Expert	Female	37	12	B.Sc.	4
Expert	Male	34	9	M.Sc.	5
Expert	Male	45	15	M.Sc.	6
Expert	Male	50	17	M.Sc.	7
Expert	Male	46	19	M.Sc.	8
Department head	Male	52	22	Ph.D.	9
Expert	Female	40	12	B.Sc.	10
Expert	Female	39	16	B.Sc.	11

Table 1. Specification of Experts

categorizes the text and creates relevant and controllable data [43]. The content used in this research was all the valid academic papers in the field of banking with the approach of using data mining. In the first step, English keywords such as “data mining in banking”, “application of data mining in bank”, “money laundering in bank + data mining”, “credit in bank + data mining”, etc., were searched in scientific databases such as Scopus, Emerald, Science Direct, Taylor, and Google Scholar, to find the most relevant articles in recent years in banking that had applied data mining in various areas of banking including money laundering, credit assessment, marketing, and so on. This research tried to include at least one paper in each application. Some areas such as “customer relationship management”, “marketing”, and “bank performance review” have been overlooked in recent years. In this cases, current paper just studies one or a few number of papers with common topic due to aim of the research that was covering various application of datamining as much as possible. A total of 17 applications were reviewed through studying 30 academic papers. Figure 1, shows the number of reviewed articles for each DM application.

Then the articles have been studied to examine what data and attributes they used. For this purpose, an Excel file was created to collect and analyze the article‘s information. In this Excel file, the “Research” sheet containing columns “row”, “title”, “year”, “authors”, “algorithm”, “sampling method”, “attributes”, “performance indicators”, “software”, “data”, “ goal”, “ result”, and “business purpose” were created to easily decide on the required data and access to them. Then, based on the extracted attributes, the initial entities were identified and then interviewed with the banking experts to confirm or reject them. The total number of 12 interviews have been conducted with 10 experts (Table 1) from various departments of a bank including customers and public banking, finance and support, banking affairs, credit affairs and corporate banking, IT and international affairs and the bank‘s IT partner. IT experts had general experience and knowledge about all areas of banking and other non-IT experts had experience and expertise in their own professional field data. Each interview lasted from 30 minutes to 2 hours.

After the final validation of the entities, the extracted attributes were linked to the entities and again interviewed with the experts to confirm the assigned attributes or to reject them. After the final validation of attributes, the main entity-relationship model was extracted with 28 entities and 423 attributes.

For applying the validated data model, one of the Iran's banks was selected for evaluation using a questionnaire and a checklist based on the model. Questionnaire divided to three part including interviewers' demography, existing recorded entities from the E-R model, and the banks business goals for applying data mining. The model's attributes were asked through a checklist and experts responded to the questionnaire and checklist.

5. Data Model for Data Analysis

5.1 Developing an Analytical Applications of Data Mining in Banking

Reviewing the articles and extracting their used information items, the extracted attributes of them were analyzed. Then, entities, attributes, and relationships between entities were designed and the relevant entity relationship model was drawn. The model initially consisted of 40 entities and 600 attributes, which were eventually reduced to 28 entities and 423 attributes after examining the model by the experts.

In Figure 2, the relationship between the entities along with the relationship degree [44], is shown. Considering the relationship degree between entities, it should be noticed that the display of the *connection degree* has been shown on the side of which entity. For example, in displaying a relationship degree between "Bank" and "Employee", the sign *..1 is displayed on the bank side, which means that one *bank* can have several *employees*, but one *employee* only belongs to one *bank*, the sign *..* in the connection between the entities of "Facility granting" and "Guarantors" which if to be read from each side of this relation has manyto-many interpretations, means that any *granted facility* can have several *guarantors*, and each *guarantor*, given his creditworthiness, can be a guarantor of *several facilities*. In fact, it should be noticed that the sign of a relationship degree has been written close to which entity, in order to interpret the relation from side of that entity. For example, in the relationship between the "Branch" and "Employee", as can be seen, the sign *...1 is located on the side of the branch, which means that any *branch* can have several *employees*, but the reverse of that isn't true, because any *employee* only belongs to one *branch*.

Due to the large number of entities and attributes, it was not possible to display all of them in a form that could be presented in this article. Figure 2 illustrates only entities. To complete Figure 2, Table 2 shows the attributes of each entity and its relationships with other entities. It is obvious that only two entities are presented in Table 2 as examples.

6. Auditing "A" Bank from the perspective of the entity-relationship data model and the availability of appropriate data for analytical applications

In order to illustrate the application of the presented data model in this article, one of the Iran's banks was selected for evaluation. Due to the lack of permission to publish the name of this bank, in the following, it is referred to as the "A" Bank. In this section, "A" Bank is first has been audited using the final entityrelationship model, Then, using the related checklist of entity-

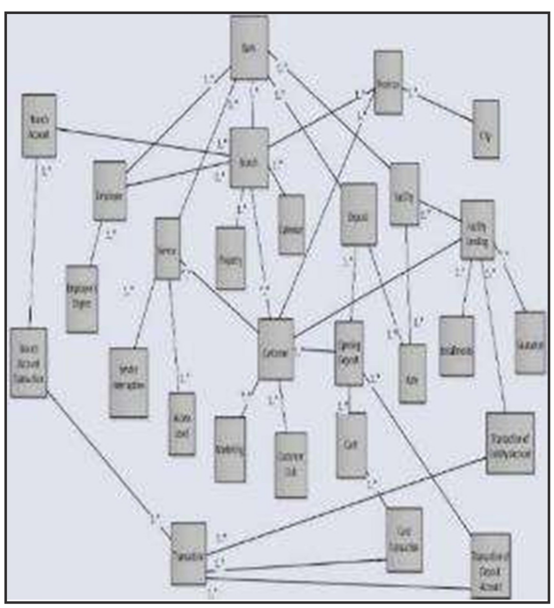

Figure 2. Final entity-relationship diagram

Entity: Bank Explanation: This entity contains information about the bank and, considering that the scope of this research cover the public and private banks, the extracted attributes is about all types of banks.				
Attributes	**Relationships with other entities**			
	One-to-One	**One-to-Many**	**Many-to-One**	**Many-to-Many**
(Economic Value Added) (Current Portfolio Value) (Size) (Attending to Financial Holding Corporation?) (Capital) Adequacy	-	Employee, Branch, Service, Deposit, Facility	-	-
Entity: Bank Explanation: This entity contains the general and identity information of the customer Attributes				
Attributes	**Relationships with other entities**			
	One-to-One	**One-to-Many**	**Many-to-One**	**Many-to-Many**
Car (Ownership) (Tax) (Payer Number) Of Dependents (Number Of Cars) (Telephone) (Gender) (Income)	-	Deposit Opening, Customer Club, Province	-	Branch, Marketing, Service
Entity: …… Explanation:…..				
	Relationships with other entities			
Attributes	**One-to-One**	**One-to-Many**	**Many-to-One**	**Many-to-Many**
.	....	...	...	...

Table 2. Entity specification

relationship model, “A” Bank has been investigated for reviewing the existence of appropriate analytical data for applications of data mining.

Using the final entity-relationship model, it was found that some entities including “Bank”, “Customer Club”, “Rate”, “Service”, “Service Interruption”, “Calendar”, and “Access Level” were not found in the Bank‘s data model, and the “Marketing” entity contained attributes other than those in the final entity-relationship model.

The Bank‘s experts stated that the “Rate” entity could be integrated into the “Deposit” and “Facility” entities, but according to the attribute obtained from the research, the need for the existence of the “Rate” entity was cleared. The auditing revealed that the bank’s portals (internet shopping through “A” bank) has not been considered in any of the existing modeled entities by the “A” Bank. At the moment of the auditing, attributes associated to the portal operations, mobile banking and internet banking have been merged into “Account Transactions” and “Card Transactions” entities. So, due to the importance of the portals for customers to use bank‘s services, it was proposed to add the “Service” and “Service Interruptions” entities to the existing entities. Regarding the “Calendar” entity, the experts stated that the calendar information can be obtained from the system, but according to the final proposed data model (Figure 2), several attributes such as “Branch Liquidity Calculation” were extracted for this entity which is very functional in banking.

The “Access Level” entity is in relation with the “Service” entity, and the explanations expressed about the “Service” entity are also true for that. Other extracted entities were modeled in existing data model of the “A” Bank.

7. Checklist for Auditing the Availability of Appropriate Data/Attributes for Analytical Applications

"A" Bank audit was conducted from the perspective of the availability of appropriate data (required attributes) for data mining using a checklist completed by the experts from various banking and IT domains. In Table 3, only two examples of entities are displayed. By providing this checklist to the "A" Bank's experts, it became clear that what entities and attributes were recorded in the bank's databases and which entities and attributes were not existed. In this way, the data gap in "A" Bank was identified for data mining analytical applications and was provided to IT professionals for using in future development of their information systems. In other words, the checklist work as a practical useful tool to guide the bank managers for improving their data maturity in order to move toward an intelligent bank because the data act as an essential infrastructure for analytical purposes.

8. Suggestions for IT Managers

Based on the findings of this research, it is suggested that banks and financial institutions analyze their existing data models based on the final proposed entity-relationship model and the proposed checklist of this research to find the existing data gaps in order to maximize the utilization of analytical applications of data mining. The proposed checklist acts as a practical guide for evaluating the required analytical data maturity. Without sufficient essential data (attributes of the model) datamining efforts will be unsuccessful. Also, according to the findings of this research, it is suggested that banks and financial institutions refer to the final entity-relationship

Entity	Attribute	Are the entity and relevant attribute register ed in "A" Bank?	Registrant information system for entity and attribute
Bank		×	Core banking
	Economic Value Added	✓	Core banking
	Current Portfolio Value	×	
	Size	×	
Customer		✓	Core banking
	Office	✓	Core banking
	Financial Expenses	×	
	Province/State	✓	Core banking
	Parent Information	✓	Core banking
	Email	✓	Core banking
	Address	✓	Core banking
	A Flag For Car Ownership	×	
	A Flag Representing Previous Employment That Lasted More Than Year	×	
	.....		

Table 3. The auditing checklist of "A" Bank

diagram in Figure 2 for development of banking information systems, in order to consider nonexisting attributes to be logged in the future and they will not face with data gaps in datamining efforts.

9. Conclusions, Limitations and Suggestions for Future Research

In this paper, according to previous research, it was found that data mining has various applications in banking that mainly involve fraud detection, credit assessment, e-banking, customer churn rate, marketing, customer lifetime value, customer satisfaction and loyalty, fraudulent accounts, phishing, customer analysis, bank performance review, customer relationship management, bank failure prediction, required cash in bank branches, money laundering, productivity analysis of the bank employees, and bank risk assessment. In order to use the data mining techniques for the abovementioned applications, banks should store properties/attributes related to the entities including the Bank, Customer, Customer Club, Employee, Employees' Degree, Branch, Branch Account, Branch Account Transaction, Transaction, Province, City, Rate, Facility, Lending Facility, Transaction of Facility Account, Installments, Deposits, Transaction of Deposit Accounts, Opening Deposit, Card, Card Transaction, Service, Service Interruption, Calendar, Property, Guarantor, Marketing, and Access Level.

In this paper, an appropriate data model was provided for banks in order to maximize their utilization of analytical applications. The proposed model developed based on the content analysis of previous research with the subject of data mining applications in banking. The entity-relationship model presented in this study shows all the entities, attributes, and relationships between them. This data model will help banks to provide the necessary data basis for analytical applications using data mining techniques. They can also use the checklist associated to this data model, which includes entities, attributes, and relationships between entities to evaluate the current status of the data. Exploring academic literature, there were existed a number of papers with various application of datamining in banking. The focus of current research for selecting previous papers was on covering various subject matter of data mining in banking to diversify the topics of DM applications as much as possible. Therefore, it may be included for example just one paper among ten existing paper in one application and this is the limitation of this study. At present, many data are collected in banks and stored in various operating systems that are raw; they can be collected in a data warehouse to perform excellent analysis for the strategic executive decisions. It is suggested to researchers to perform a similar research based on the results of this research in order to design a data warehouse architecture. The second suggestion is to perform such a research for other industries and organizations for example healthcare, insurance, and so on.

References

[1] Shu, W., Strassmann, P. A. (2005). Does information technology provide banks with profit? *Information & Management*, 42 (5) 781-787. doi:10.1016/j.im.2003.06.007

[2] Agarwal, R., Dhar, V., (2014). Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for is Research, *Information Systems Research*, 25 (3) 443-448. https://doi.org/10.1287/isre.2014.0546

[3] Turban, E., Sharda, R., Delen, D. (2014). *Decision support and business intelligence systems*. Essex: Pearson.

[4] Moro, S., Cortez, P., Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, 42 (3) 1314-1324. doi:10.1016/j.eswa.2014.09.024

[5] Bhasin, M. L. (2006). Data Mining: A Competitive Tool in the Banking and Retail Industries. Banking and finance.

[6] Pulakkazhy, S., Balan, R. (2013). Data Mining In Banking and Its Applications-A Review. *Journal of Computer Science*, 9 (10) 1252-1259. doi:10.3844/jcssp.2013.1252.1259

[7] Ngai, E., Hu, Y., Wong, Y., Chen, Y., Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50 (3) 559-569. doi:10.1016/j.dss.2010.08.006

[8] Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J. C. (2011). Data mining for credit card fraud: *A comparative study. Decision Support Systems*, 50 (3) 602-613. doi:10.1016/j.dss.2010.08.008.

[9] Wei, W., Li, J., Cao, L., Ou, Y., Chen, J. (2012). Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, 16 (4) 449-475. doi:10.1007/s11280-012-0178-0

[10] Yap, B. W., Ong, S. H., Husain, N. H. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications*, 38 (10) 13274-13283. doi:10.1016/j.eswa.2011.04.147

[11] Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: *A market comparative study. Decision Support Systems*, 37 (4) 543- 558. https://doi.org/10.1016/S0167- 9236(03)00086-1

[12] Gulati, R., Goswami, A., Kumar, S., (2018). What drives credit risk in the Indian banking industry? An empirical investigation, *Economic Systems*, https://doi.org/10.1016/j.ecosys.2018.08.004.

[13] Ali, G., Aritürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with*

Applications, 41 (17) 7889-7903. doi:10.1016/j.eswa.2014.2014.06.018

[14] Curko, K., Bach, M. P., Radonic, G. (2007). Business Intelligence and Business Process Management in Banking Operations. 2007 29th International Conference on Information Technology Interfaces. doi:10.1109/iti.2007.4283744.

[15] Azadeh, A., Saberi, M., Jiryaei, Z. (2012). An intelligent decision support system for forecasting and optimization of complex personnel attributes in a large bank. Expert Systems with Applications, 39 (16) 12358-12370. doi:10.1016/j.eswa.2012.04.056

[16] Hughes, A. M. (1996). The Complete Database Marketer: Secondgeneration Strategies and Techniques for Tapping the Power of Your Customer Database, McGraw Hill.

[17] Abdou, H. A., Tsafack, M., Ntim, C. G., Baker, R. (2016). Predicting Creditworthiness in Retail Banking with Limited Scoring Data. *SSRN Electronic Journal*. doi:10.2139/ssrn.2756746.

[18] Srivastava, U., Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks, *Procedia Computer Science*, Vol. 50, p. 643 – 652, https://doi.org/10.1016/j.procs.2015.04.098.

[19] Mandalaa, Narindra., G. N., Nawangpalupia, Badra., C., Praktikto, Rian., F. (2012). Assessing Credit Risk: an Application of Data Mining in a Rural Bank, *Procedia Economics and Finance*, Vol. 4, p. 406 – 412, https://doi.org/10.1016/S2212-5671(12)00355-3

[20] Pérez-Martín, A., Pérez-Torregrosa, A., Vaca, M. (2018). Big Data techniques to measure credit banking risk in home equity loans, *Journal of Business Research*, Vol. 89, p. 448-454, https://doi.org/10.1016/j.jbusres.2018.02.008.

[21] Batmaz, I., Danisoglu, S., Yazici, C., Kartal-Koç, E. (2017). A data mining application to deposit pricing: *Main determinants and prediction models*, *Applied Soft Computing*, Vol. 60, p. 808 819, https://doi.org/10.1016/j.asoc.2017.07.047.

[22] Tavana, M., Patnaik, S. (2018). Recent Developments in Data Science and Business Analytics: Proceedings of the International Conference on Data Science and Business Analytics (ICDSBA- 2017). Cham: Springer International Publishing.

[23] Li, S., Yen, D. C., Lu, W., Wang, C. (2012). Identifying the signs of fraudulent accounts using data mining techniques. *Computers in Human Behavior*, 28 (3) 1002-1013. doi:10.1016/j.chb.2012.01.002.

[24] Cao, D. K., Do, P. (2012). Applying Data Mining in Money Laundering Detection for the Vietnamese Banking Industry. Intelligent Information and Database Systems Lecture Notes in Computer Science, 207-216. doi:10.1007/978-3-642-28490-8_22.

[25] Aburrous, M. R., Hossain, A., Dahal, K., Thabatah, F. (2009). Modelling Intelligent Phishing Detection System for E-banking Using Fuzzy Data Mining. 2009 International Conference on CyberWorlds. doi:10.1109/cw.2009.43

[26] Liébana-Cabanillas, F., Nogueras, R., Herrera, L., Guillén, A. (2013). Analysing user trust in electronic banking using data mining methods. *Expert Systems with Applications*, 40 (14) 5439-5447. doi:10.1016/j.eswa.2013.03.010

[27] Miguéis, V. L., Camanho, A. S., Borges, J. (2017). Predicting direct marketing response in banking: comparison of class imbalance methods. *Service Business.* doi:10.1007/s11628-016-0332-3

[28] Ekinci, Y., Uray, N., Ülengin, F. (2014). A customer lifetime value model for the banking industry: a guide to marketing actions. *European Journal of Marketing*, 48 (3/4) 761- 784. doi:10.1108/ejm-12-2011-0714

[29] Hsieh, N. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, 27 (4) 623-633. doi:10.1016/j.eswa.2004.06.007

[30] Au, W., Chan, K. (2003). Mining fuzzy association rules in a bank-account database. *IEEE Transactions on Fuzzy Systems,* 11 (2) 238-248. doi:10.1109/tfuzz.2003.809901

[31] Lin, S., Shiue, Y., Chen, S., Cheng, H. (2009). Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks. *Expert Systems with Applications*, 36 (9) 11543-11551. doi:10.1016/j.eswa.2009.03.029

[32] Carmona, P., Climent, F., Momparler, A. (2018). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach, *International Review of Economics & Finance*, https://doi.org/10.1016/j.iref.2018.03.008.

[33] Climent, F., Momparler, A., Carmona, P. (2018). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach, *Journal of Business Research*, https://doi.org/10.1016/j.jbusres.2018.11.015.

[34] Zhao, H., Sinha, A. P., Ge, W. (2009). Effects of feature construction on classification performance: An empirical study in bank failure prediction. *Expert Systems with Applications,* 36 (2) 2633-2644. doi:10.1016/j.eswa.2008.01.053

[35] Costa, G., Folino, F., Locane, A., Manco, G., Ortale, R. (2007). Data Mining for Effective Risk Analysis in a Bank Intelligence Scenario. 2007 IEEE 23rd International Conference on Data Engineering Workshop. doi:10.1109/icdew.2007.4401083

[36] Baghbani, G., Eskandari, F. (2017). Calculating the required cash in bank branches: a Bayesian-data mining approach. *Neural Computing and Applications*. doi:10.1007/s00521-017-2888-9

[37] Ponniah, P. (2007). Data modeling fundamentals: a

practical guide for IT professionals. Hoboken, NJ: Wiley-Interscience.

[38] Ambler, S. W. (2003). Agile database techniques: effective strategies for the agile software developer. Hoboken, NJ: Wiley.

[39]. Silberschatz, A. (2013). Database System Concepts. S.L. Mcgraw-Hill Education.

[40] Bagui, S. S., Earp, R. (2011). Database design using entity-relationship diagrams. Boca Raton, FL: Auerbach.

[41] Sekaran, U., Bougie, J. R. (2016). Research methods for business: a skill-building approach. Chichester: Wiley.

[42] Ghosh, B. N. (2002). Designing social research. Leeds: Wisdom House.

[43] Weber, R. (1990). Basic Content Analysis. doi:10.4135/9781412983488.

[44] Eren, C. (2008). Nested bitemporal relational data model.