

# A Study of Data Requirements for Data Mining Applications in Banking

Mina Ranjbarfard, Shahideh Ahmadi

Alzahra University

Iran

{m.ranjbarfard@alzahra.ac.ir} {shahideh.ahmadi@gmail.com}



Journal of Digital  
Information Management

**ABSTRACT:** *There are many studies that have applied data mining to banking. However, the lack of proper data mounts a serious obstacle to the employment of data mining techniques by banks. This paper examines previous data mining research in the field of banking to extract all served entities and attributes required for analytical purposes, categorize these attributes and ultimately present a data model for analysis. After analyzing a wide range of data mining applications in banking, 28 entities with 423 attributes were identified and the final proposed entity-relationship model was drawn. Also, a checklist was provided based on the model for auditing data gap in banks and applied to a real case. The results of this paper can be seen as a supportive tool for improving bank's business intelligence maturity from the data perspective and enabling managers for analyzing data requirement of information systems.*

## Subject Categories and Descriptors

[H.2.8 Database Applications]; Data mining; [D.3.3 Language Constructs and Features]; Data types and structures

**General Terms:** Data Mining, Banking Data, Data Analysis

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

**Received:** 19 November 2019, Revised 27 February 2020, Accepted 9 March 2020

**Review Metrics:** Review Scale: 0/6, Review Score: 5.1, Inter-reviewer Consistency: 68%

**DOI:** 10.6025/jdim/2020/18/3/109-117

## 1. Introduction

Banking is a fertile industry for new information systems and technologies [1] including business intelligence and decision support systems. As a key concept of machine learning and big data [2], business intelligence includes architectures, tools, databases and methods of data analysis to support decision making for business executives [3]. Banking domains such as branch efficiency, e-banking, customer segmentation and retention, to mention a few, provide an extraordinary context for broad application of business intelligence concepts and methods such as data mining, data warehouses and decision support systems [4], [5], [6]. Data mining methods of business intelligence have been applied to improve banking operations such as fraud detection [7], [8], [9], credit assessment [10], [11], [12], customer churn prediction, etc. [13]. Today, banks have become increasingly aware of the value of their customer related data. They need to transform their daily recording of transaction data for complex analysis including risk management, customer relationships, profit and production channels, market valuation, operational efficiency, etc. [14]. Failure in these areas yields undesirable outputs such as loss of customers, property damage, loss of credit and heavy fines [6]. However, one may argue that the required data attributes are not consolidated and organized to perform data analysis. In order to create knowledge in these domains, the relevant data gathered by the information systems is necessary and the lack of suitable data poses a serious obstacle to employment of data mining techniques for the banks. However, scant research in the field of banking has investigated items required for exploiting data mining in the banks. Basically, the question is "what is suitable data model in banks to maximize analytical applications?" and

“how can one assess the richness of existing data in banks with the aim of applying data mining techniques?” Therefore, we decided to examine various research about data mining applications in banking and extract their served data attributes and relationship in the form of a data model (in the form of entity-relationship model) so that banks can obtain the necessary data required to make plans for the development of their information systems to keep pace with the latest progress in intelligent banking. In this context, this research seeks to identify various applications of data mining in banking as well as relevant data attributes reside in banking information systems to provide a comprehensive data model for promoting data mining and business analysis applications. This model helps IT managers of banks to consider the analytical requirements in the long-term planning of information systems development so that required data for data mining applications and analysis capabilities will be provided in their bank. Based on the proposed data model, we developed a checklist for auditing data gap in real cases. Also we audited Bank “A” as a sample for showing the application of the proposed data model and its linked checklist. The results of this paper can be seen as a supportive tool to improve bank’s business intelligence maturity based on data perspective.

## 2. Theoretical Foundations and Literature Review

### 2.1. Application of Data Mining in Banking

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

Banking industry is increasingly discovering the importance of customer information, which encompasses a large volume of personal information, demographic information, transactional data, credit card usage patterns and so forth. 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 including customer information, transaction details, risk profiles, credit card details, collateral details, compliance, anti-money laundering information and Swift messages. Thousands of decisions are made on a daily basis including issues related to credits, fraud, investment, money laundering and illegal financial. A manager needs to obtain different reports to make critical decisions by adopting various banking system tools [6]. Using data mining techniques to analyze patterns and trends, bank managers can predict how customers react to new rates or which customers are likely to accept new products.

There are several applications of data mining in the banking

system. For example, the use of logistic regression, Classification and Regression Tree (CART) and Cascade Correlation Neural Network (CCNN) to build knowledge-based scoring models [17; 18] reveal how big data analytics is being successfully recruited in the banking sector to explore facets such as the spending pattern of customers, channel usages, customer segmentation and profiling, product cross selling, sentiment, feedback analysis, security and fraud management. Credit risk represents the main trend in banking. There is also a growing interest in bankruptcy and fraud prediction. Customer retention seems to be associated, though weakly, with targeting, justifying bank offers to reduce customer churn [4].

Credit risk assessment for secured loans is a critical operation [19] that is aimed at identifying necessary factors to assess credit risk, and generate a new decision tree model based on C 5.0 method. It seeks to reduce the number of non-performing loans [20], develop Monte Carlo experiments based on known techniques and algorithms and implement a linear mixed model (LMM) as a new contribution to calculate the credit risk of financial companies [21].

Studied determinants of deposit pricing using various methods such as multivariate adaptive regression splines, support vector regression, artificial neural networks, classification and regression trees and random forest highlight the importance of customer- and account-specific characteristics in the determination of deposit rates. Also, it has been shown that depositors who forge a multi-faceted and long-term relationship with a 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. Their model predicts liquidity risk by identifying its triggering factors. [5] examined support vector machines (SVM) and random forests with logit regression to detect fraud using real data obtained from international credit card payment transactions. Random forests excelled other techniques in terms of performance, though logit regression and SVMs had desirable performance as well. [23] studied the Taiwan Bank using Bayesian classification and association rule to identify the signs of fraudulent accounts and the patterns of fraudulent transactions. Detection rules were developed based on identified signs and applied to the design of a fraudulent account detection system. Applying clustering techniques to money transferring data in Vietnamese banks, [24] proposed some approaches to money laundering detection techniques. They presented a system for detecting money laundering using the CLOPE algorithm.

[25] studied the Bank of England and proposed a model based on fuzzy logic and data mining algorithms to characterize the factors of e-banking phishing website. The techniques are evaluated by classifying diverse phishing types and defining six criteria to attack e-banking phishing websites in a layered structure. In a case study on Spanish Bank, [26] reviewed various methods and techniques

of determining the important variables required for financial institutions to predict probable levels of trust among e-banking users. To do so, they employed the most recent advances in machine learning and soft-computing, including a new selection operator for multi-objective genetic algorithms. The proposed method yielded optimum results in terms of optimization and the highest punctuation by the experts. [13] in a case study of European private banks, proposed a dynamic customer churn prediction framework for generating training data from customer records to predict customer churn within multiple horizons by standard classifiers. [27], in a case study of Portage Bank, proposed a data mining response model supported by random forests to reinforce the definition of target customers. They compared the performance of an underdamping method (the Easy Ensemble algorithm) with an oversampling method (the Synthetic Minority Oversampling technique) besides exploring the importance of attribute features included in the response model. Random forests supported by an underdamping algorithm demonstrated an excellent prediction performance compared to other techniques. [28] in a 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) were compared to select the best performing forecasting tool to predict the potential CLV. The LSE-based linear regression model was selected due to its more desirable performance. The proposed model contained not only profit and cost-oriented indicators, but also other indicators rarely used in the literature. [29] looked at the account data of customers and their credit card transactions in Taiwan Bank to discover interesting patterns in the data with the goal of providing clues about incentives that a company can offer as a marketing strategy to its customers. A two-stage behavioral scoring model was presented with a cascade involving self-organizing 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 improve its services and retain customers by discovering customer's hidden patterns. [31] applied the particle swarm optimization (PSO) to obtain suitable parameter settings for the support vector machine (SVM) and decision tree (DT), selecting a subset of useful features in order to predict bank performance. According to the experimental results, their proposed approaches significantly enhanced the accuracy of classification. [32] predicted bank failure in the U.S. banking sector using extreme gradient boosting (XGBoost) by drawing on the annual series of 30 financial ratios in 156 national commercial banks in the United States. [33] applied Extreme Gradient Boosting to 25 annual series data of financial ratio to identify a set of key indicators that help predict and thwart bank failure in the Eurozone banking sector. In a study on the Bank of America, [34] used four data mining methods including logistic regression, decision tree, neural network, and k-nearest neighbor to predict bank failure. [15] in a study on 102 branches of a large private bank, presented an intelligent integrated decision support system (DSS) by

merging data mining tools (RST, ANN, MLP, GA, and CVTT), DEA and K-Means for the prediction and optimization of complex personnel efficiency. [35] proposed a data warehousing architecture for effective risk analysis in banking. Moreover, they presented a hash-based technique for data reconciliation. In a study on an Iranian bank, [36] classified 18 branches of a certain bank to determine cash sufficiency in bank's branches. Using hierarchical clustering and Bayesian hierarchical clustering in identical clusters, they estimated the amounts of cash received and used by each branch through neural network.

## 2.2. Data Model and Entity

Data model is a method for defining real-world information requirements and make them understandable to the organization's shareholders. In addition, data modeling enables database specialists to draw on information requirements for implementation of a computer database system in order to support the organization. Therefore, a data model provides a major tool for communication with users as well as database system design for developers [37]. The entity type is like the concept of a class in an 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 include customer, order and tax [38]. An entity is represented by a set of attributes. Attributes are descriptive properties possessed by each entity, including name, last name, address and telephone for the customer entity [39]. The ER<sup>1</sup> model is one of the best known tools for database logic design. This model is understandable by non-professionals and can be easily visualized. Besides, their entities and their relationship are visual, which can display the user's information requirements in a natural manner. Database systems are typically modeled using an entity-relationship diagram (model) as a design for the stored real data [40].

## 3. Research Methodology

This is an applied research in terms of its purpose [41]. It looks for an entity-relationship data model to identify the current data gap in banks and financial institutions. With respect to the implementation process, the present research is qualitative [42]. We have adopted the content analysis method, which is used to categorize texts and create relevant and controllable data [43]. In this research, we reviewed available valid academic papers in the field of banking with the aim of using data mining. In the first step, keywords including "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 including Scopus, Emerald, Science Direct, Taylor, and Google Scholar to find the most relevant articles in recent years that had applied data mining to various areas of banking such as money laundering, credit assessment, marketing, and so forth. We

<sup>1</sup> Entity Relationship Model

have tried to include at least one paper for each application. Some areas such as “customer relationship management”, “marketing”, and “bank performance review” have been overlooked in recent years. In such cases, this paper only studies a few papers with common themes because the research goal is to cover various application of data mining. A total of 17 applications were reviewed by studying 30 academic papers. Figure 1 shows the number of articles reviewed for each DM application.

The articles were reviewed to examine the type of data attributes used. For this purpose, an Excel file was created to collect and analyze articles’ information. In this file, the “Research” sheet was created with columns such as “row”, “title”, “year”, “authors”, “algorithm”, “sampling method”, “attributes”, “performance indicators”, “software”, “data”, “ goals”, “ results”, and “business purposes” to determine and access the required data. Then, based on the extracted attributes, the initial entities were identified. We then interviewed the banking experts to confirm or reject the entities and their relevant attributes. A total

of 12 interviews were conducted with 10 experts (Table 1) from various departments of a bank including customers and public banking, finance and support, banking affairs, credit affairs, corporate banking, IT and international affairs and the bank’s IT partner. IT experts generally had experience and knowledge about all domains of banking and non-IT experts had expertise in their own professional field. Each interview lasted from 30 min to 2 h.

After the final validation of the entities, attributes and their relationships, the main entity-relationship model was extracted with 28 entities and 423 attributes.

To apply the validated data model, one of the Iranian banks was selected. The evaluation was conducted using a questionnaire and a checklist based on the model. The questionnaire was divided into three parts including demographic information, recorded entities and relevant attributes from the entity-relationship model, and business goals of banks to be used for the data mining. The model attributes were presented through a checklist and experts responded to the questionnaire and checklist.

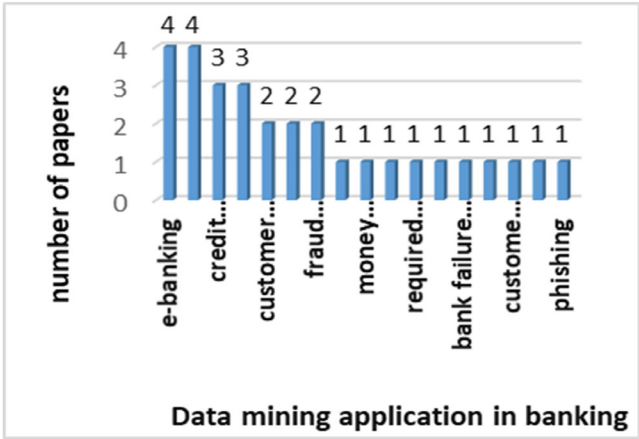


Figure 1. Frequency diagram of data mining applications

Organizational Position	Sex	Age	Experience	Education	Row
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. Specifications of Experts



## 4. Data Analysis

### 4.1. Developing the Data Model for Analytical Application of Data Mining in Banks

We reviewed the articles and extracted their information items to analyze the extracted attributes. Then, entities, attributes, and inter-entities relationships were designed and the relevant entity relationship model was drawn. The initial model consisted of 40 entities and 600 attributes, which were eventually reduced to 28 entities and 423 attributes after the model review by experts.

Figure 2, shows the relationship between entities as well as the strength of such relationships. [44] The strength of relationship between entities is displayed as the *connection degree* on the side of each entity. For example, to display the strength of relationship between “Bank” and “Employee”, sign \*.1 is shown on the bank side, meaning that a *bank* can have multiple *employees*, but an *employee* only belongs to one *bank*. Also, the sign \*.\* regarding the association between the entities of “granting facilities” and “guarantors,” which can have several interpretations if read from each side of this relation, means

that a *granted facility* can have several *guarantors*, and each *guarantor*, based on his creditworthiness, can be a guarantor of *several facilities*. In fact, it is worth noting that the sign of the strength of relationship is written near each entity to allow interpreting the relation from the side. For example, regarding the relationship between “Branch” and “Employee”, the sign \*.1 is shown on the side of the branch, meaning that a *branch* can have several *employees*, but the reverse is not true, for an *employee* only belongs to one *branch*.

Since there were a large number of entities and attributes, it was not possible to display all of them in this paper. In Figure 2, only entities have been shown. as a supplement to Figure 2, Table 2 shows the attributes of each entity and its connections to other entities. We have shown only two entities in Table 2 as examples.

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

To audit a specific bank in terms of the availability of required data (attributes) for data mining, a checklist was developed based on the entity-relationship model. In Table

<b>Entity: Customer</b> <b>Note: This entity contains general and identity information of customers</b>				
Attributes	Relationships with other entities			
	One-to-One	One-to-Many	Many -to- One	Many -to-Many
(Car Ownershi) (Tax Payer) (Number of Dependents) (Number of Cars)(Telephone) (Gender) (Income) .....	-	Deposit Opening, Customer Club, Province	-	Branch, Marketing, Service
<b>Entity: .....</b> <b>Explanation:.....</b>				
Attributes	Relationships with other entities			
	One-to-One	One-to-Many	Many -to- One	Many -to-Many
.	....	...	...	...

<b>Entity: Bank</b> <b>Notes: This entity contains information about the bank</b>				
Attributes	Relationships with other entities			
	One-to-One	One-to-Many	Many -to- One	Many -to-Many
(Economic Value Added) (Current Portfolio Value) (Size) (Equity) (Capital Strength)	-	Employee, Branch, Service, Deposit, Facility	-	-

Table 2. Entity specifications



3, only a short part of this checklist is presented. By giving this checklist to a bank's experts, one can find what entities and attributes were recorded in bank databases and which entities and attributes were missing. In this way, the data gap in considered bank will be determined. IT professionals can use the data gap report for future development of information systems. In other words, the checklist served as a practical tool and guide for bank managers to improve their data maturity and move towards intelligent banking, considering that data lays the ground for analytical purposes.

### 4.3. Model Validation

To validate the proposed entity relationship model, we conducted some interviews with experts in several rounds.

Firstly, after identifying the required entities from the literature, we conducted some sequential interviews to find whether the list of entities and their related attributed were corrected or not. Then we modified the entities and related attributes according to expert's comments after each interview and went to the next interview until the list of identified entities and their related attributed was confirmed by the experts.

Secondly, we checked the relationship among entities through some sequential interviews with experts. After correcting the relationships according to the expert's comments, the final entity-relationship model was developed which is presented in figure 2.

### 5. Case Study

To illustrate the application of the data model presented in this article, one of the Iranian banks was selected for evaluation. Since we are not permitted to publish the name of this bank, it is referred to as Bank "A" in this paper. In this section, Bank "A" is audited using the final entity-relationship model. To do this, based on the relevant checklist of entity-relationship model, Bank "A" is investigated to check the existence of any attributes of data model.

Using the final entity-relationship model, we found that some entities including "Bank", "Customer Club", "Rate", "Service", "Service Interruption", "Calendar", and "Access Level" were absent in the Bank's data model, and the "Marketing" entity comprised attributes other than those mentioned in the final entity-relationship model.

According to the Bank's experts, the "Rate" entity could be integrated into the "Deposit" and "Facility" entities, but the attributes obtained in this research suggest the importance of having a "Rate" entity. The results of auditing revealed that the bank's portals (Internet shopping in Bank "A") have not been considered in existing entities modeled for the Bank "A". During auditing, attributes associated with portal operations, mobile banking and Internet banking were merged into "Account Transactions" and "Card Transactions" entities. Hence, given the importance

of portals for customers to use bank services, we proposed the addition of Service" and "Service Interruptions" entities to existing entities. With respect to the "Calendar" entity, the experts declared that the calendar information could 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 of significant importance for the banking.

The "Access Level" entity is related to the "Service" entity, and points mentioned about the "Service" entity also hold true for that. Other extracted entities were modeled in the existing data model of Bank "A".

### 6. Conclusions

This paper, consistent with previous research, found that data mining has diverse applications in banking, including 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, cash required in bank branches, money laundering, productivity analysis of the bank's employees, and bank risk assessment. To apply data mining techniques for the above-mentioned purposes, banks need to store properties/attributes related to the entities including the Bank, Customers, Customer Club, Employees, Employees' Degree, Branch, Branch Account, Branch Account Transaction, Transaction, Province, City, Rate, Facilities, Lending Facilities, 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.

The entity-relationship model presented in this study exhibits all the entities, attributes, and their relationships. The proposed model was developed based on the content analysis of previous research on data mining applications in banking.

Based on the findings of this research, banks and financial institutions are recommended to analyze their existing data models based on the proposed entity-relationship model. They can also use the checklist linked to this data model, which includes entities, attributes, and inter-entity relationships to evaluate data status. The proposed checklist can be served to discover the existing data gaps in order to maximize the use of analytical applications of data mining in banking. This checklist acts as a practical guide for evaluating the analytical data maturity. In the absence of sufficient essential data (attributes of the model) the data mining efforts will be unsuccessful. Also, according to the results, it is suggested that banks and financial institutions refer to the entity-relationship diagram in Figure 2 for the development of banking information systems to consider non-existing

attributes in the future and avoid data gaps in data mining efforts.

A review of literature suggested that there are a number of papers about various applications of datamining in banking. In this research, we selected papers that covered various domains of data mining in banking to diversify the topics of DM applications. Therefore, we may have included just one paper out of ten existing papers in one application, which is one of the limitations of this study. At present, a large volume of raw data is collected in banks and stored in various operating systems. Such data can be collected in a data warehouse to perform analyses related to the strategic executive decisions. Hence, it is suggested that researchers perform similar studies based on the results of this research to design a data warehouse architecture. Moreover, it is suggested to perform such a research in other industries and organizations such as healthcare, insurance, etc.

## 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](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., Arýtürk, 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.2014.06.018
- [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, Pages 643 – 652, <https://doi.org/10.1016/j.procs.2015.04.098>.
- [19] Narindra Mandalaa, G. N., Badra Nawangpalupia, C., Rian Praktikto, F. (2012). Assessing Credit Risk: an Application of Data Mining in a Rural Bank, *Procedia Economics and Finance*, Vol. 4, Pages 406 – 412, [https://doi.org/10.1016/S2212-5671\(12\)00355-3](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, Pages 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 model, *Applied Soft Computing*, Vol. 60, Pages 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.