Data Mining Methodologies in the Banking Domain: A Systematic Literature Review

Veronika Plotnikova, Marlon Dumas, Fredrik P. Milani

University of Tartu, Institute of Computer Science, J. Liivi 2, 50409 Tartu, Estonia
name.surname@ut.ee

Abstract. Data mining and advanced analytics methods and techniques usage in research and in business settings have increased exponentially over the last decade. Development and implementation of complex Big Data and advanced analytics projects requires well-defined methodology and processes. However, it remains unclear for what purposes and how data mining methodologies are used in practice and across different industry domains. This paper addresses the need and provides survey in the field of data mining and advanced data analytics methodologies, focusing on their application in the banking domain. By means of systematic literature review we have identified 102 articles and analyzed them in view of addressing three research questions: for what purposes data mining methodologies are used in the banking domain? How are they applied ("as-is" vs adapted)? And what are the goals of adaptations? We have identified that a dominant pattern in the banking industry is to use data mining methodologies as-is in order to tackle Customer Relationship Management and Risk Management business problems. However, we have also identified various adaptations of data mining methodologies in the banking domain, and noticed that the number of adaptations is steadily growing. The main adaptation scenarios comprise technology-centric aspects (scalability), business-centric aspects (actionability) and human-centric aspects (mitigating discriminatory effects).

Keywords: Data Mining · Banking · Literature Review

1 Introduction

The use of data mining methodologies have gained significant adoption in business settings, in particular in the financial services sector [1]. However, little is known about what and how data mining methodologies are applied. There are studies that surveyed data mining techniques and applications across domains, yet, they focused on data mining process artefacts and outcomes (eg.[2]), but not on end-to-end process methodology. There are some studies that have surveyed data mining methodologies in hospitality [3], accounting [4], education [5], and manufacturing [6] industries, but no comprehensive studies have been conducted on financial companies. In particular, studies in banking domain were so far narrow in scope - either addressed only specific data mining techniques, typically in

connection with concrete business problem or product domain (eg. credit cards [7]), or tackled the technique in combination with required software toolset [8]. Data mining process methodology in this research was not addressed.

Given this gap, we investigate the application of data mining methodologies in the banking domain[1]. This is achieved by tackling the following research questions: for what purposes data mining methodologies are used in the banking domain? (RQ1), how are they applied ("as-is" vs adapted)? (RQ2), and what are the goals of adaptations? (RQ3).

The research questions are addressed by the means of a systematic literature review (SLR). As part of SLR, existing studies have been categorized by deriving taxonomy, and examined in depth by analyzing typical data mining methodologies application scenarios. The paper provides two distinct contributions: (1) it identifies and classifies data mining methodologies application scenarios and business problems addressed in banking industry settings, (2) it examines data mining methodologies adaptations, documenting associated reasons, goals and benefits. In doing so, the paper identifies gaps in 'de-facto' standard data mining methodologies that manifest themselves when applied in banking. Further, it provides evidence and insights to built upon further research activities with respect to data mining frameworks applications in banking domain.

The work is structured as follows. Section 2 provides the background while Section 3 presents the research design. The findings are presented in Section 4 while Section 5 concludes.

2 Background

The section provides a brief overview of data mining concept, existing data mining methodologies and their evolution.

Data mining methodologies can be defined as a set of rules, processes, algorithms that are designed to generate actionable insights, extract patterns, and identify relationships from large data sets [9]. As such, data mining methods commonly involve extraction, processing, and modeling data by means of methods and techniques.

The data mining methods are commonly represented as a high level process [10, 11] that defines a set of activities and tasks, inputs and outputs required, accompanied with guidelines on how to perform the steps [10]. The foundations for structured data mining were first proposed by [12–14] with the introduction of Knowledge Discovery in Databases (KDD). This approach consists of nine steps. The first concerns learning the application domain by which is meant understanding the domain and identifying the goals of data mining. The second step focuses on creating the dataset while the third works with data cleaning

[1] We use the term *banking domain* to refer to: (1) traditional businesses providing universal banking and insurance products and services (eg. lending, transactions, capital markets, asset management, etc.) to all types of clientele (private, corporate, financial institutions and firms), and (2) niche players, disruptors (FinTech, monoline banks etc.) specialized in specific banking, insurance products and services

and processing. The fourth step, data reduction and projection, concerns finding useful features to represent the data. In the fifth step, the target outcome is defined while in the sixth step, the methods and models to use on the dataset, with consideration to the objectives, are selected. In the seventh step, the work of mining the data is performed followed by the eight step where the results are interpreted and finally, are used as basis for decisions (ninth step).

The KDD approach gained traction in industrial and academic settings [11, 15], and it was also used as basis for refinements aiming to address specific gaps. However, such approaches received limited attention [11, 15] with the exception of SEMMA (Sample, Explore, Modify, Model and Assess). The latter has been widely adopted due to its incorporation into SAS data mining tool [16].

An industry-driven methodology called Cross-Industry Standard Process for Data Mining (CRISP-DM) was introduced in 2000 as an alternative to KDD [11]. CRISP-DM is considered as 'de-facto' standard for data mining methodology and commonly used as a reference framework by which other methodologies are benchmarked against [10]. While CRISP-DM builds upon KDD, it consists of six phases that are executed in iterations [11]. The iterative executions of CRISP-DM stands as the most distinguishing feature when compared to KDD. CRISP-DM, much like KDD, aims at providing practitioners with guidelines to perform data mining on large data sets and designed to be domain-agnostic [10]. As such, it is widely used by industry and research communities [11].

CRISP-DM has six phases with a total of 24 tasks and outputs. The first phase is to understand the business domain, the project objectives, and converting business requirements into data mining problem definition. In the second phase, the objective is to gain an initial understanding of the data. The third phase focuses on data preparation while in the fourth phase various modelling techniques are selected and applied. In the fifth phase, the models are evaluated to ensure that they can achieve the objectives. In the final (sixth) phase, the models are deployed and results organized, presented, and distributed. Similarly to KDD, CRISP-DM has been used as basis for new data mining approaches which largely addressed deployment, use of insights [17] or project management and organizational factors [18]. CRISP-DM has also been modified to specific domains such as Industrial Engineering [19] and Software Engineering [20].

3 Research Design

The main research objective of this paper is to study how data mining methodologies are applied in the banking domain. We apply systematic literature review (SLR) method as it ensures trustworthy, rigorous, and auditable methodology, as well as supports synthesis of existing evidence, identification of research gaps, and provides framework to appropriately position new research activities [21]. Our SLR followed the guidelines proposed by [21].

To formulate the research questions, we started from the traditional set of "W" questions, specifically "Why?", "What?" and "How?". The "Why" question led us to RQ1 (for what purposes are data mining methodologies used in the

banking domain?). We then raised the "What" question ("What data mining methodologies are used in the banking domain?"), but discarded this question after a preliminary analysis - we found that all major data mining methodologies (e.g. CRISP-DM, SEMMA, etc.) are used in this domain and there are little insights to be derived from analyzing this question further. Next, we raised the "How?" question, which led us to RQ2 (are data mining methodologies in the banking domain used "as-is" or are they adapted?). An initial exploration of this question led us to the preliminary conclusion that indeed data mining methodologies are sometimes adapted, which in turn led us to pose a third research question: With what goals are data mining methodologies adapted for the banking domain (RQ3)?

According to the guidelines for conducting SLR [21] we derived and validated search terms and strings, identified types of literature, selected electronic databases, and defined the screening procedures.

The search string were derived from the research questions and included the terms "data mining" and "data analytics" as these are often used interchangeably. The terms "methodology", "framework" and "banking" were added resulting in the search string being defined as ("data mining methodology") OR ("data mining framework") OR ("data analytics methodology") OR ("data analytics framework") AND ("banking"). Validation of the search string according to [22], led to adding the search string of ("CRISP-DM") OR ("SEMMA") OR ("ASUM") AND ("banking") in order to capture case study papers. The search strings were applied to Scopus, Web of Science, and Google Scholar databases. Multidisciplinary indexed/non-indexed electronic databases were selected to ensure wide data sources coverage, and to include studies from both academic (peer reviewed) and practitioners communities ("grey" literature). Specifically, our "grey" literature search covered industry reports, white papers, technical reports, and research works not indexed by Scopus or Web of Science.

Based on the SLR best practices [21], we designed a multi-step screening procedures (relevancy and quality) with associated set of *Screening Criteria* (exclusion and inclusion criteria), and *Scoring System*. The exclusion criteria served to eliminate studies in languages other than English, duplicating texts, as well as publications shorter than 6 pages, or the ones not accessible (by University subscriptions). Papers that passed all exclusion criteria were retained and assessed according to relevance criteria. Each paper was considered relevant if it was: (1) about data mining approach within the banking domain, and (2) introduced or described data mining methodology/framework or modification of existing approaches. Finally, quality screening was conducted for full texts evaluation. For that we developed a *Scoring Metrics* as proposed in [22]. Papers were given the score of 3 if all steps of the data mining process were clearly presented and explained. Further, to merit a score of 3, the paper must have also presented proposal on usage, application, or deployment of solution in organization' s business process(es) and IT/IS system, and/or discuss prototype or full solution implementation. If description of some process steps were missing, but without impacting the holistic view and understanding of the work performed,

the paper was given a score of 2. Only papers scoring "2" or "3" were included in the final primary studies corpus.

The initial number of studies retrieved amounted to 693 of which 167 were academic and 526 "grey" literature. Having performed the screening based on exclusion criteria, 509 studies remained and were subject to relevance screening. 141 papers were finally identified as relevant and moved into quality assessment phase, and 41 peer-reviewed papers and 61 studies from "grey" literature received a score of 2 or higher. By means of SLR we identified primary texts corpus with 102 relevant studies. Figure 1 below exhibits yearly published research numbers with the breakdown by "peer-reviewed" and "grey" literature starting from 1997.

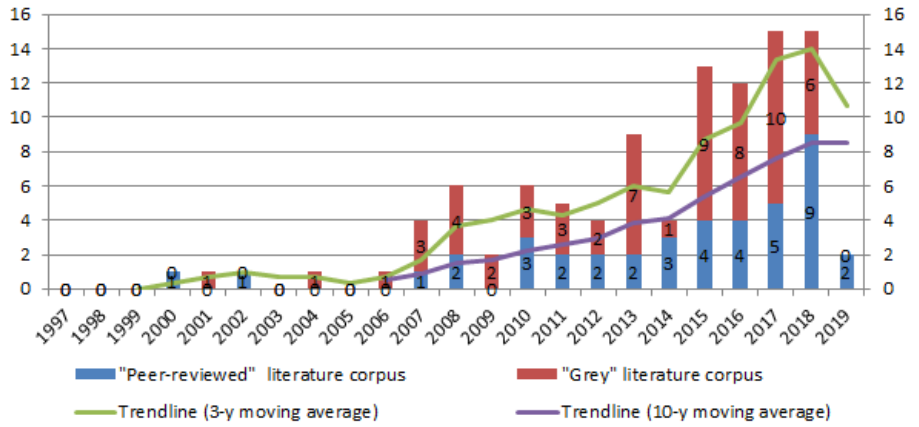

Fig. 1. SLR derived texts corpus - data mining methodologies peer-reviewed research and "grey" literature for period 1997-2019 (no. of publications).

Temporal analysis of texts corpus resulted in two observations. Firstly, we note that research on application of data mining methodologies within the banking domain began more than a decade ago - in 2007. Research efforts made prior to 2007 were infrequent and irregular, with 3-4 years gap periods between publications. Secondly, we note that research on data mining methodologies within banking domain has grown since 2007, an observation supported by the 3-year and 10-year constructed mean trendlines. In particular, we also note that the number of publications have roughly tripled over the last decade, hitting all-time high in 2018 with 22 texts released.

4 Findings and Discussion

In this section, we present results of publications analysis, address the research questions and discuss threats to validity.

RQ1 - For what purposes are data mining methodologies used in the banking domain? In-depth analysis of text corpus revealed that data mining methodologies are predominantly being employed in the banking domain for two main purposes - customer-oriented and risk-oriented (see Figure 2a below).

We identified 47 customer-oriented studies which address various aspects related to customer behavior modelling. A typical example is profiling according to usage pattern of different digital channels, [24][2] authors profiled Internet bank users, while [25] focuses on patterns of electronic transactions based on demographic and behavioural features. In the field of Customer Relationship Management (CRM), the most common business problem analyzed relate to identifying and predicting customers who are likely to churn [26], customer loyalty and retention [27], customer segmentation [28], and customer value identification [29]. Further, smart and improved customer targeting in sales campaigns [30] and improved targeting and customer prioritization decision support are also popular business problem [31]. A few studies consider efficiency aspects of bank's infrastructure such as Automated Teller Machines (ATMs) and branch networks (eg. [32]).

The second most commonly analyzed area is Risk Management, predominantly, credit risk. We identified 34 studies that focus on modelling tasks for supporting a variety of risk management processes including credit risk scoring and default prediction([33]), prediction of financial distress [34], and credit decisions for private and corporate customers (especially, small and medium enterprises as in [35]). Further, identification and prevention of fraud behavior [36] and AML (anti-money laundering) risks [37] are addressed as well. Finally, other risk management topics, such as market risk, as well as asset management [38], trading strategies [39], overall economic analysis and predictions [53] are also addressed.

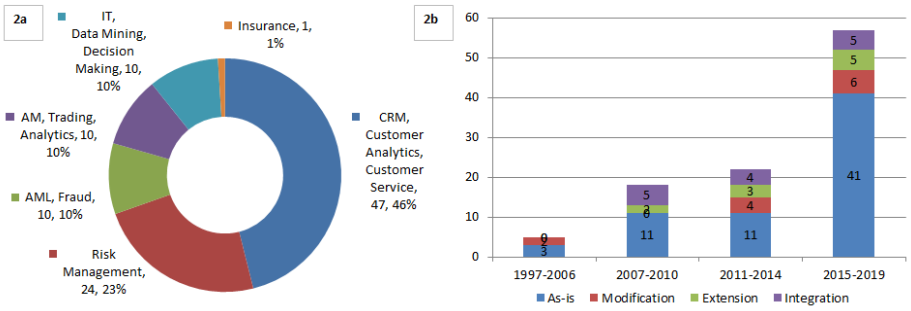

Fig. 2. Applications of data mining methodologies in banking: 2a) breakdown by purposes; 2b) breakdown by adaptation paradigms

[2] Due to space limitation, examples of key texts are presented throughout the analysis. All texts corpus is available at https://figshare.com/articles/MasterList xlsx/8206604

RQ2 - How are data mining methodologies applied ("as-is" vs adapted)? The second research questions addresses the extent to which data mining methodologies are used "as-is" versus adapted. Our review identified two distinct paradigms on how data mining methodologies are applied. The first is "as-is" where the data mining methodologies are applied as stipulated. The second is with "adaptations", i.e., methodologies are modified by introducing various changes to the standard process model when applied. Furthermore, our review led us to identify three distinct adaptation scenarios namely "Modification", "Extension", and "Integration":

Scenario "Modification" - introduces specialized sub-tasks and deliverables in order to address a specific use cases or business problems. Modifications typically concentrate on granular adjustments to the methodology at the level of sub-phases, tasks or deliverables within the existing CRISP-DM or KDD stages.

Scenario "Extension" - primarily proposes significant extensions to CRISP-DM resulting in either fully-scaled and integrated data mining solutions, data mining frameworks as a component or tool for automated IS systems or adapted to specialized environments. Adaptations where extensions have been made elicit and explicitly presents various artefacts in the form of system and model architectures, process views, workflows, and implementation aspects. Key benefits achieved are deployment, implementation and leveraging of data mining solutions as integral components of IS systems and business processes. Also, data mining process methodology is substantially changed and extended in all key phases to accommodate new Big Data technologies, tools and environments ([47, 53]).

Scenario "Integration" - 'Integration' primarily concentrates on either combining CRISP-DM with data mining methodologies originated from other domains (e.g. Business Information Management, Business Process Management, BI [58]), adjusting to specific organizational aspects [62], and discrimination-awareness with respect to customers [56]. Adaptations in the form of integration typically introduces various types of ontologies and ontology-based tools, business processes, business information, and BI-driven framework elements. Key benefits are improved at the deployment phase, improved usage of data and discovered knowledge, higher business processes effectiveness and efficiency. Key gap filled in is lack of CRISP-DM integration with other organizational and domain frameworks.

We also noted that publications discussing "as-is" implementations have grown strongly but at the same time, adaptations are also gaining ground (as exhibited in Figure 2b). Further, there is balanced development and distribution of the research among "Modification", Extension" and "Integration" paradigms. We can hypothesize that existing reference methodologies do not accommodate and support increasing complexity of data mining projects and IS/IT infrastructure, as well as banking domain specific requirements and as such need to be adapted.

RQ3 - What are the goals of adaptations? We address the third research question by analyzing each of adaptation scenarios in depth.

Modification This adaptation scenario was identified in 12 publications where modifications overwhelmingly consist of specific case studies. However, the major differentiating point compared to "as-is" case studies is clearly the presence of specific adjustments towards standard data mining process methodologies. Yet, the proposed modifications and their purposes do not go beyond traditional CRISP-DM phases. They are granular, specific and executed on tasks, sub-tasks, and at the level of deliverables. This is in clear contrast to "extensions" where one of the key proposals are new phases, such as including a new IS/IT systems implementation and integration phase. Also, with modifications, authors describe potential business applications and deployment scenarios at a conceptual level, but typically do not report or present real implementations to the IS/IT systems and business processes. Further, in the context of bank-

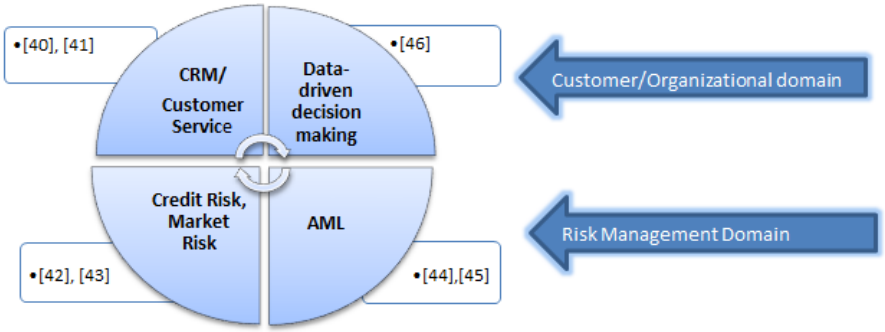

Fig. 3. Data Mining Methodologies in Banking - 'Modification' scenario example texts mapping to business problems

ing domain, this research subcategory can be classified with respect to business problems addressed (presented in the Figure 3 [3].)

Extension "Extension" scenario was identified in 10 publications and we noted that it was executed for the two major purposes:

1. To implement fully scaled, integrated data mining solution and regular, repeatable knowledge discovery process - address model, algorithm deployment, implementation design (including architecture, workflows and corresponding IS integration). Also, complementary goal is to address changes to business process to incorporate data mining into organization activities
2. To implement complex, specifically designed systems and integrated business applications with data mining model/solution as component or tool.

[3] Due to space limitations, two most cited texts references are presented if number of texts per category exceed two, all texts corpus is available at https://figshare.com/articles/MasterList xlsx/8206604

Typically, this adaptation is also oriented towards Big Data specifics, and is complemented by proposed artefacts such as Big Data architectures, system models, workflows, and data flows.

We also conclude that the first purpose focuses on implementation of specific data mining models and associated frameworks and processes. For example, apart from classification model and evaluation framework, [47] proposes a knowledge-rich financial risk management process while [48] introduces framework for machine-learning audits. [49] presented data mining-based solution for AML implemented as a tool with respective IS architecture and investigative process. [50] focused on combined data mining concept introducing multiple data sources, methods and features, all incorporated in the real-time prototyped solution. [51] focused on actionable data mining by presenting post-processing data mining framework which enables automated actions generation. In the similar vein, [52] presented large-scale data mining framework extended to incorporate social media data including adaptions to parallel processing. The major benefit achieved by these adaptations, apart from resolved business problem or research gap, is the usefulness of results produced in the decision-making process.

In contrast, the second purpose concentrates on design of complex, multi-component information systems and architectures. For instance, [53] have constructed a framework that considers socio-economic data, its processing methods, a new data life-cycle model, and presented an architecture for Big Data systems to integrate, process and analyze data for forecasting purposes. [54] proposed refinements of reference data mining methodology to address Big Data analytics, applications prototyping and its evaluation, project management and results communication. Finally, [55] proposed cross-border market monitoring and surveillance system with 3 subsystem components, system and data flows. In this research, authors discuss and present useful architectures, algorithms and tool sets in addition to methods and techniques which alone are not sufficient to create deployable systems and tools. The key benefits provided are broad context enabling practical implementations of complex, integrated data mining solutions. The specific list of studies mapped to each of the given purposes along with key artefacts is presented in Figure 4 below.

Integration Integration of data mining methodologies were found in in 14 publications. Our analysis shows that these adaptations are at the highest abstraction level and typically executed with the goals to (1) introduce discrimination-awareness in data mining, (2) integrate/combine with other organizational frameworks, and (3) integrate/combine with other well-known frameworks, process methodologies and concepts. Example list of studies with artefacts is presented in Figure 4[4] and further discussed.

Discrimination-aware data mining (DADM), as proposed by [56], includes tool support for "correct" decision process. The major benefit is increased correctness and usefulness of results in the decision-making process, monitoring, avoidance of discrimination and transparency.

[4] All texts corpus with complete mappings is available at https://figshare.com/articles/MasterList xlsx/8206604

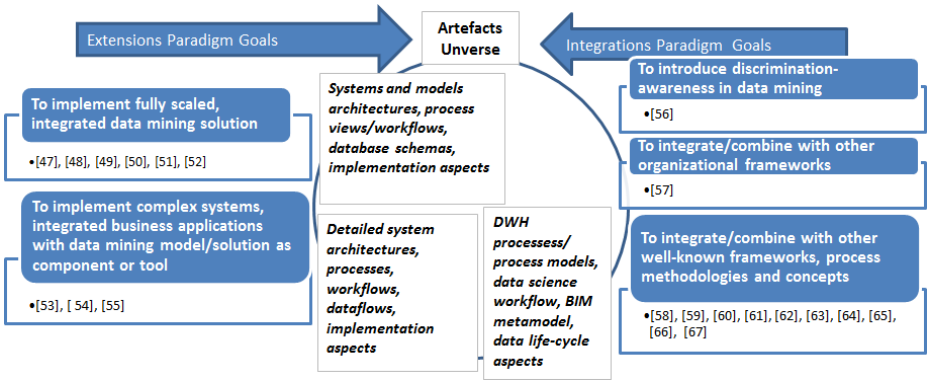

Fig. 4. Data Mining Methodologies in Banking - 'Extension' and 'Integration' scenarios adaptation goals, their artefacts and example texts mapping

[57] author combined data mining methodology with organizational context to instill and improve data-driven decision-making. Further, [58] integrated data mining with business process frameworks and models (also proposed by [59]). [60] integrated data mining with BIM (Business Information Modelling) while [61] merged data mining with BI. All with the purpose to improve usage of data, business processes effectiveness and deployment of data mining solutions. These works are complemented by number of publications [62, 63] specifically tackling actionability of data mining results, which aim to reduce likelihood of data mining project producing high quality knowledge with limited or no business benefit. Authors propose shift to domain-driven data mining paradigm by integrating such new key component as domain intelligence, human-machine cooperation, in-depth mining, actionability enhancement, and iterative refinement process. Emphasis on data-mining business requirements, model sharing and resuse from business user perspective is also tackled by introducing ontology-based data mining model management approach [64]. Identical problems are addressed from organizational point of view by [65], which focused on Big Data Analytics governance framework. Finally, number of innovative research papers focused on integrating data mining with technical concepts and frameworks from other domains, for example, relational (symbolic) data mining methods [66] and game theory [67].

To summarize, from "extension" and "integration" research we have identified three important banking domain specific factors, which require adjustments of existing data mining process frameworks and models. Firstly, potential discrimination in the context of credit decision-making requires financial services companies to adapt data mining to achieve transparency. Secondly, large number of accumulated data and associated complex IS/IT architectures, require to adapt data mining process to address complex data mining models deployment patterns and implement them as component of complex systems and business applications. Thirdly, actionability of data mining results, adaptation of analyt-

ics outcomes to end, business-user needs are of utmost importance to achieve business value realization. We can hypothesize that in banking domain as the leading adopter of data mining solutions with significant investments, failures of realizing full business value of data mining projects are more explicit and observable and need to be addressed.

This study has inherent threats to validity and limitations associated with the selected research method (SLR). The validity threats include incompleteness of search results (internal validity[5]) and general publication bias (external validity[6]). We have mitigated internal validity by strictly adhering to inclusion criteria, and performing significant validation procedures. With respect to external validity, we conducted trial searches to ensure validity of search strings and proper identification of potential papers. Our initial publications harvest size reached almost 700 texts originated from indexed peer-review research and "grey" literature thus mitigating external validity risk. Further, the key limitation of the SLR method for this study is that banking industry internal practices are not frequently disclosed in academic literature. We mitigated the negative impact by inclusion of "grey" literature where reporting on existing industry practices by professionals is common.

5 Conclusion

In this study we have examined data mining methodologies usage in the banking domain. By means of Systematic Literature Review we have identified 102 relevant studies of peer-reviewed and "grey" literature which have been evaluated in depth to address three research questions: for what purposes data mining methodologies are used in the banking domain? (RQ1), how are they applied ("as-is" vs adapted)? (RQ2), and what are the goals of adaptations? (RQ3).

Tackling RQ1 (For what purposes?) we have discovered that data mining methodologies are applied regularly since 2007 and their usage has tripled. Further, data mining in financial services domain is primarily used for two main purposes - to address Customer Relationship Management and Risk Management related business problems.

Answering RQ2 (How?), we have identified that over the last decade data mining methodologies have been primarily applied "as-is" without modifications. Yet, we have also discovered emerging and persistent trend of using data mining methodologies in banking with adaptations. Further, we have distinguished three adaptations scenarios ranging from granular modifications on tasks, subtask and deliverables level and ending up with merging standard data mining methodologies with other frameworks.

Addressing RQ3 (What are the adaptations goals?), we have examined the adaptation objectives, banking domain specific factors behind such adaptations,

[5] The internal validity stems from subjective screening and rating of studies when applying relevancy and quality criteria

[6] The threats to external validity relate to the extent by which the results can be generalized beyond the scope of this study

and as a result have identified three such aspects. Firstly, discriminatory awareness and transparent decision-making (human-centric aspect) require data mining process adaptation. Secondly, actionability of data mining results (business-centric aspect) plays a central role in the banking domain. Thirdly, we have also identified that standard data mining methodologies lack deployment and implementation aspects (technology-centric aspects) required to scale and transform data mining models into software products and components integrated into Big Data Architectures. Therefore, adaptations are used to integrate data mining models and solutions in complex IT/IS systems and business processes of the banking industry. This study highlighted the needs and established ground for future work to develop refinements of existing data mining methodologies for the banking domain which would address three abovementioned concerns.

References

1. Forbes Homepage, https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/4cf12a2139a1, last accessed 2019/05/26
2. Liao, S. H., Chu, P. H., Hsiao, P. Y.: Data mining techniques and applications A decade review from 2000 to 2011. Expert systems with applications, **39**(12), 11303–11311 (2012)
3. Mariani, M., Baggio, R., Fuchs, M., Hoepken, W.: Business intelligence and big data in hospitality and tourism: a systematic literature review. International Journal of Contemporary Hospitality Management, **30**(12), 3514-3554 (2018)
4. Amani, F., Fadlalla, A.: Data mining applications in accounting: A review of the literature and organizing framework. International Journal of Accounting Information Systems, **24**, 32-58 (2017)
5. Murnion, P., Helfert, M.: A framework for decision support for learning management systems. In: 10th European Conference on e-Learning ECEL-2011. Brighton, UK (2011)
6. Zhuming, Bi, Cochran, D.: Big data analytics with applications. Journal of Management Analytics, **1**(4), 249–265 (2014)
7. Wongchinsri, P., Kuratach, W.: A survey - data mining frameworks in credit card processing. In: 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 1–6. IEEE, Chiang Mai, Thailand (2016)
8. Hassani, H., Huang, X., Silva, E.: Digitalisation and big data mining in banking. Big Data and Cognitive Computing, **2**(18), 1-14 (2018)
9. Morabito, V.: The future of digital business innovation: Trends and practices. Springer, Switzerland (2016)
10. Mariscal, G., Marban, O., Fernandez, C.: A survey of data mining and knowledge discovery process models and methodologies. Knowledge Engeneering Review, **25**(2), 137-166 (2010)
11. Marban, O., Mariscal, G., Segovia, J.: A data mining and knowledge discovery process model. Data Mining and Knowledge Discovery in Real Life Applications, edited by P. Julio and K. Adem, pp. 438-453, Paris, I-Tech, Vienna, Austria (2009)
12. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Magazine, **17**(3), 37-54 (1996a)

13. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P.: The KDD process for extracting useful knowledge from volumes of data. Communications ACM, **39**(11), 27-34 (1996b)
14. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P.: Knowledge discovery and data mining: Towards a unifying framework. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), pp. 82-88, Portland, Oregon, USA (1996c)
15. Kurgan, L. A., Muslek, P.: A survey of knowledge discovery and data mining process models. Knowledge Engineering Review, **21**(1), 1-24 (2006)
16. SAS Institute: Data Mining Using SAS Enterprise MinerTM: A Case Study Approach. SAS Institute Inc., Cary, 1166NC (2013)
17. Cios, K. J., Kurgan, L. A.: Trends in data mining and knowledge discovery. Advanced techniques in knowledge discovery and data mining, pp. 1-26, Springer (2005)
18. Moyle, S., Jorge, A.: Ramsys - a methodology for supporting rapid remote collaborative data mining projects. In: ECML/PKDD01 Workshop: Integrating Aspects of Data Mining, Decision Support and Meta-learning (IDDM-2001) (2001)
19. Solarte, J.: A Proposed Data Mining Methodology and its Application to Industrial Engineering. PhD Thesis, University of Tennessee (2002)
20. Marban, O., Segovia, J., Menasalvas, E., Fernandez-Baizan, C.: Toward data mining engineering: A software engineering approach. Information systems, **34**(1), 87-107 (2009)
21. Kitchenham, B.: Procedures for performing systematic reviews. Keele University Technical Report 1038TR/SE-0401, ISSN:1353-7776; NICTA Technical Report 0400011T.1, 1-28 (2004)
22. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report No. EBSE-2007-01 (2007)
23. Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., Khalil, M.: Lessons from applying the systematic literature review process within the software engineering domain. Journal of Systems and Software, **80**(4), 571-583 (2007)
24. Mansingh, G., Rao, L., Osei-Bryson, K. M., Mills, A.: Application of a Data Mining Process Model: A Case Study-Profiling Internet Banking Users in Jamaica. In AMCIS, 439 (2010)
25. Etaiwi, W., Biltawi, M., Naymat, G.: Evaluation of classification algorithms for banking customers behavior under apache spark data processing system. Procedia computer science, **113**, 559–564 (2017)
26. Kumar, D. A., Ravi, V.: Predicting credit card customer churn in banks using data mining. International Journal of Data Analysis Techniques and Strategies, **1**(1), 4–28 (2008)
27. Bahari, T. F., Elayidom, M. S.: An efficient CRM-data mining framework for the prediction of customer behaviour. Procedia computer science, **46**, 725–731 (2015)
28. Tsiptsis, K. K., Chorianopoulos, A.: Data mining techniques in CRM: inside customer segmentation. John Wiley and Sons, UK (2011)
29. Moeini, M., Alizadeh, S. H.: Proposing a New Model for Determining the Customer Value Using RFM Model and its Developments (Case Study on the Alborz Insurance Company). Journal of Engineering and Applied Sciences, **100**(4), 828–836 (2016)
30. Neysiani, B. S., Soltani, N., Ghezelbash, S.: A framework for improving find best marketing targets using a hybrid genetic algorithm and neural networks. In: 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), pp. 733-738. IEEE (2015)

31. Ghosh, S., Hazra, A., Choudhury, B., Biswas, P., Nag, A.: A Comparative Study to the Bank Market Prediction. In: International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 259–268. Springer, Cham (2018)
32. Met, I., Tunali, G., Erko, A., Tanrikulu, S., Dolgun, M. O.: Branch Efficiency and Location Forecasting: Application of Ziraat Bank. Journal of Applied Finance and Banking, **7**(4), 1–13 (2017)
33. Khemakhem, S., Ben Said, F., Boujelbene, Y.: Credit risk assessment for unbalanced datasets based on data mining, artificial neural network and support vector machines. Journal of Modelling in Management, **13**(4), 932–951 (2018)
34. Geng, R., Bose, I., Chen, X.: Prediction of financial distress: An empirical study of listed Chinese companies using data mining. European Journal of Operational Research, **241**(1), 236–247 (2015)
35. Gulsoy, N., Kulluk, S.: A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **9**(3), e1299 (2019)
36. Adeyiga, J. A., Ezike, J. O. J., Omotosho, A., Amakulor, W.: A neural network based model for detecting irregularities in e-Banking transactions. African Journal of Computer and ICTs, **4**(2), 7–14 (2011)
37. Colladon, A. F., Remondi, E.: Using social network analysis to prevent money laundering. Expert Systems with Applications, **67**, 49–58 (2017)
38. Liu, X., Ye, Q.: The different impacts of news-driven and self-initiated search volume on stock prices. Information and Management, **53**(8), 997–1005 (2016)
39. Al-Radaideh, Q. A., Assaf, A. A.,Alnagi, E.: Predicting stock prices using data mining techniques. In: The International Arab Conference on Information Technology (ACIT2013) (2013).
40. Smith, K. A., Willis, R. J., Brooks, M.: An analysis of customer retention and insurance claim patterns using data mining: A case study. Journal of the operational research society, **51**(5), 532–541 (2000)
41. Karimi-Majd, A. M., Mahootchi, M.: A new data mining methodology for generating new service ideas. Information Systems and e-Business Management, **13**(3), 421–443 (2015)
42. Montiel, J., Bifet, A., Abdessalem, T.: Predicting over-indebtedness on batch and streaming data. In: 2017 IEEE International Conference on Big Data (Big Data), pp. 1504–1513. IEEE (2017)
43. Rajan, M.: Credit Scoring Process using Banking Detailed Data Store. International Journal of Applied Information Systems (IJAIS), **8**(6), 13–20 (2015)
44. Luo, X.: Suspicious transaction detection for anti-money laundering. International Journal of Security and Its Applications, **8**(2), 157–166 (2014)
45. Resta, M.: VaRSOM: A Tool to Monitor Markets' Stability Based on Value at Risk and SelfOrganizing Maps. Intelligent Systems in Accounting, Finance and Management, **23**(1-2), 47–64 (2016)
46. Kaddouri, A.: Why Human Expertise is Critical for Data Mining. International Journal of Computer and Information Technology, **2**(1), 99–108 (2013)
47. Peng, Y., Wang, G., Kou, G., Shi, Y.: An empirical study of classification algorithm evaluation for financial risk prediction. Applied Soft Computing, **11**(2), 2906–2915 (2011)
48. Clark, A.: The Machine Learning Audit - CRISP-DM Framework. ISACA, **1** (2018)
49. Le Khac, N.A., Markos, S., Kechadi, M.T.: A data mining-based solution for detecting suspicious money laundering cases in an investment bank. In: Second International Conference on Advances in Databases, Knowledge, and Data Applications, pp. 235–240. IEEE (2010)

50. Sridevi, P., Reddy, N.: Informative Knowledge Discovery using Multiple Data Sources, Multiple Features and Multiple Data Mining Techniques. IOSR Journal of Engineering, **3**1, 20–25 (2013)
51. Yang, Q.: Post-processing data mining models for actionability. In: Data mining for business applications, pp. 11–30. Springer, Boston, MA (2009)
52. Yuan, H., Lau, R. Y., Xu, W., Pan, Z., Wong, M.: Mining Individuals Behavior Patterns from Social Media for Enhancing Online Credit Scoring. In: 22nd Pacific Conference on Information Systems (PACIS) Proceedings, 163, Japan (2018)
53. Blazquez, D., Domenech, J.: Big Data sources and methods for social and economic analyses. Technological Forecasting and Social Change, **130**, 99–113 (2018)
54. Ange, S., Lozano-Argel, S. I., Montoya-Munera, E. N., Ospina-Arango, J. D., Tabares-Betancur, M. S.: Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data and Analytics Projects. In: International Conference on Knowledge Management in Organizations, pp. 613–624. Springer, Cham (2018)
55. Diaz, D., Theodoulidis, B., Abioye, E.: Cross-Border Challenges in Financial Markets Monitoring and Surveillance: A Case Study of Customer-Driven Service Value Networks. In: 2012 Annual SRII Global Conference, pp. 146–157. IEEE (2012)
56. Berendt, B., Preibusch, S.: Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence. Artificial Intelligence and Law, **22**(2), 175–209 (2014)
57. Debuse, J. C. W.: Extending data mining methodologies to encompass organizational factors. Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research, **24**(2), 183–190 (2007)
58. Pivk, A., Vasilecas, O., Kalibatiene, D., Rupnik, R.: On approach for the implementation of data mining to business process optimisation in commercial companies. Technological and economic development of economy, **19**(2), 237–256 (2013)
59. Lessmann, S., Listiani, M., Vo, S.: Decision Support in Car Leasing: a Forecasting Model for Residual Value Estimation. In: 31st International Conference on Information System (ICIS) Proceedings, 17, St. Louise (2010)
60. Priebe, T., Markus, S.: Business information modeling: A methodology for data-intensive projects, data science and big data governance. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2056–2065. IEEE (2015)
61. Balkan, S., Goul, M.: A portfolio theoretic approach to administering advanced analytics: The case of multi-stage campaign management. In: 44th Hawaii International Conference on System Sciences, pp. 1–10. IEEE (2011)
62. Cao, L., Zhang, C.: The evolution of KDD: Towards domain-driven data mining. International Journal of Pattern Recognition and Artificial Intelligence, **21**(04), 677–692 (2007)
63. Cao, L.: Domain-driven data mining: Challenges and prospects. IEEE Transactions on Knowledge and Data Engineering **22**(6), 755–769 (2010)
64. Li, Y., Thomas, M.A., Osei-Bryson, K.M.: Ontology-based data mining model management for self-service knowledge discovery. Information Systems Frontiers, **19**(4), 925–943 (2017)
65. Lawler, J., Joseph, A.: Big Data Analytics Methodology in the Financial Industry. Information Systems Education Journal, **15**(4), 38–51 (2017)
66. Kovalerchuk, B., Vityaev, E.: Symbolic methodology for numeric data mining. Intelligent Data Analysis, **12**(2), 165–188 (2008)
67. Qin, Z., Wan, T., Dong, Y., Du, Y.: Evolutionary collective behavior decomposition model for time series data mining. Applied Soft Computing, **26**, 368–377 (2015)