

Title	Digitalisation and Big Data Mining in Banking
Type	Article
URL	https://ualresearchonline.arts.ac.uk/id/eprint/13228/
Date	2018
Citation	Hassani, H. and Huang, X. and Silva, E.S. (2018) Digitalisation and Big Data Mining in Banking. Big Data and Cognitive Computing, 2. pp. 1-18.
Creators	Hassani, H. and Huang, X. and Silva, E.S.

Usage Guidelines

Please refer to usage guidelines at <http://ualresearchonline.arts.ac.uk/policies.html> or alternatively contact ualresearchonline@arts.ac.uk.

License: Creative Commons Attribution Non-commercial No Derivatives

Unless otherwise stated, copyright owned by the author



Article

Digitalisation and Big Data Mining in Banking

Hossein Hassani ^{1,*} , Xu Huang ² and Emmanuel Silva ³

¹ Research Institute of Energy Management and Planning, University of Tehran, Tehran 1417466191, Iran

² Faculty of Business and Law, De Montfort University, Leicester LE1 9BH, UK; xu.huang@dmu.ac.uk

³ Fashion Business School, London College of Fashion, University of the Arts London, London WC1V 7EY, UK; e.silva@fashion.arts.ac.uk

* Correspondence: hassani.stat@gmail.com

Received: 27 June 2018; Accepted: 17 July 2018; Published: 20 July 2018



Abstract: Banking as a data intensive subject has been progressing continuously under the promoting influences of the era of big data. Exploring the advanced big data analytic tools like Data Mining (DM) techniques is key for the banking sector, which aims to reveal valuable information from the overwhelming volume of data and achieve better strategic management and customer satisfaction. In order to provide sound direction for the future research and development, a comprehensive and most up to date review of the current research status of DM in banking will be extremely beneficial. Since existing reviews only cover the applications until 2013, this paper aims to fill this research gap and presents the significant progressions and most recent DM implementations in banking post 2013. By collecting and analyzing the trends of research focus, data resources, technological aids, and data analytical tools, this paper contributes to bringing valuable insights with regard to the future developments of both DM and the banking sector along with a comprehensive one stop reference table. Moreover, we identify the key obstacles and present a summary for all interested parties that are facing the challenges of big data.

Keywords: big data analytics; data mining; banking; survey

1. Introduction

The era of big data came along with both big opportunities and challenges, almost all science subjects are experiencing overflowing information at unpredictable volume and speeds [1]. As such, revealing the hidden information in big data via Data Mining (DM) techniques has become an emerging trend and ultimate objective for a wide range of studies [2–5]. As a data intensive subject, banking has been a popular implementation field for researchers with DM skills over the past decades of the information science revolution. Banks have acknowledged that knowledge instead of financial resources is the new biggest asset [6]. Moreover, the development and popularization of e-banking and mobile banking adds to the exponential growth of real time banking information. These continuous developments and the rapidly increasing availability of big data make mastering relevant big data analytics tools one of the most crucial tasks for the banking sector.

Following a comprehensive investigation of existing literature, to the best of our knowledge, only two review papers focused on the DM applications in banking [7,8] and both covered a number of DM implementations before 2013. For such a rapidly developing subject that progresses on a daily basis, it is important to provide researchers and interested parties with the most up to date status of DM and banking collaborations. As such, we thoroughly review the DM applications in banking, especially for the recent years, post 2013. It is noteworthy that we will not repeat the contents covered in [7,8], but instead focus on the most recently developed DM applications in the banking sector. This paper aims to serve as the most up to date one stop directory guide for relevant researchers and apprise them of the evolution of big data analytics in banking with an outlook for future research.

Having summarized the recent applications, big data in banking have been exploited for improving customer satisfaction, marketing and optimizing strategic management. Specifically, the recent applications collected in this paper targeted mainly four topics: security and fraud detection, risk management and investment banking, customer relationship management (CRM) and other advanced supports. Moreover, this paper contributes by gathering and analyzing the information of DM techniques, software, and data resources (among other factors). As such, the contents of this review ensures its novelty and meaningful contributions to the existing gap in academic literature.

The remainder of this paper is organized such that the methodology is presented in Section 2; the value creation of DM in banking by topics along with a summarized reference table are reviewed in Section 3; the key techniques and software trends are briefly summarized in Section 4; Section 5 finally concludes and proposes directions and challenges for future research.

2. Methodology

We adopt the methodology presented in [9] considering its sophisticated research design, which is clearly demonstrated to the reader via Figure 1.

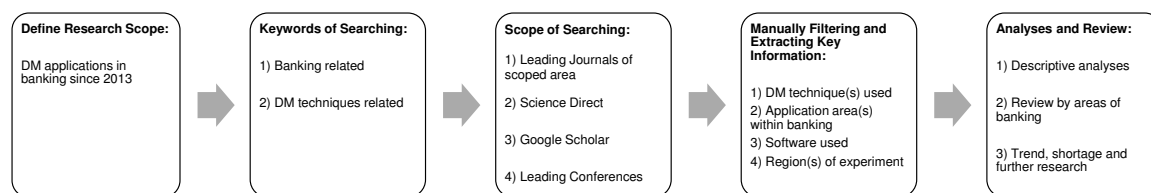


Figure 1. Research framework for DM applications in banking.

Firstly, as clarified in the last section, the research scope is defined as the DM applications in banking post 2013. The search process follows the usual approach by defining a range of keywords, where we employed the significant terms for both banking and DM techniques, including banking, fraud detection, credit card, credit scoring, risk management, deposit, mortgage, debit, loan, CRM, bank marketing; and data mining, clustering, text mining, classification and other specific DM technique terms. It is noteworthy that the pairwise searching approach is conducted while ensuring at least one term from each aspect is represented. Moreover, the scope of the search focuses solely on reliable academic resources from leading journals (for example, Expert Systems with Applications, Decision Support Systems, Data Mining and Knowledge Discovery) and leading conferences (e.g., IEEE series, and Knowledge Discovery and Data Mining). To this end, there are over 100 applications identified to be manually reviewed and summarized in the following section.

3. Value Creation of DM in Banking by Topics

Having reviewed over 100 recent DM applications in banking post 2013, it can be generally concluded that the banking sector mainly adopts DM techniques for the following purposes:

Security and fraud detection: Big secondary data like transaction records are monitored and analysed to enhance banking security and distinguish the unusual behavior and patterns indicating fraud, phishing, or money laundering (among others).

Risk management and investment banking: Analysis of in-house credit card data freely accessible for banks enables credit scoring and credit granting which form part of the most popular tools for risk management and investment evaluation.

CRM: DM techniques have been widely applied in banking for marketing and customer relationship management related purposes such as customer profiling, customer segmentation, and cross/up selling. These help the banking sector to have a better understanding of their customers, predict customer behaviour, accurately target potential customers and further improve customer satisfaction with a strategic service design.

Other advanced supports: A few less mainstream applications focus on branching strategy, and efficiency and performance evaluation, which can significantly assist in achieving strategic branch locating and expansion plans.

In what follows, we briefly review the collected publications with respect to the areas of interests. Moreover, to serve as the one stop reference directory for all recent DM applications in banking post 2013, a summary table can be found in Table 1, where the literature are grouped by means of value creation, along with detailed information like data resources/regions and DM techniques adopted. Note that most implementations are applied to more than one DM technique, and some applications did not clarify the specifics due to confidential restrictions.

3.1. Security and Fraud Detection

In order to maintain the high standards of security amidst the overwhelming flow of big banking data and the rapidly growing scale and complexity of cyber crimes, researchers have been exploring advanced DM techniques for effectively identifying unusual fraud behaviour. Note that an existing review that targeted credit card processing can be found in [10]. From an internal perspective, the survey data of bank employees in India are collected in [11] to analyse their perceptions with regard to fraud.

Many researchers worked with transaction data, seeking better approaches to distinguish between patterns from genuine behavior with higher efficiency and accuracy [12–25]. Among these, Wei et al. [12] proposed a framework named i-Alertor for major Australian banks; a semi-supervised decision support system named BankSealer was proposed in [14] for an Italian bank; authors in [15] proposed a hybrid DM method to predict network intrusions and detect fraud activities; FraudMiner model that integrated frequent itemset mining was introduced in [16] and verified with the data set from UCSD DM contest 2009; a comparative study [17] addressed the ensemble approach to build classifiers; in terms of a recent advancement in FraudMiner, the authors in [18] introduced the LINGO clustering technique [26] for the pattern matching process, and this enhancement helped maintain a satisfying performance in terms of accuracy while further reducing the false alarm rate; Behera and Panigrahi [19,20] demonstrated the hybrid approach for credit card fraud detection by combining Fuzzy Clustering and NN techniques, and achieved over 93% accuracy with the dataset generated in [27]; APATE was proposed in [21] for automated fraud detection within a large credit card issuer in Belgium; both Luhn's and Hunt's algorithms were employed in [22] for proposing a novel system of credit card fraud detection; the authors in [23] illustrated the use of DM techniques on customer data to add a higher level of authentication to banking processes for real time fraud detection; a hybrid approach combining genetic algorithm and NN was proposed in [24] for Greek companies in the banking sector; a framework named FDiBC was developed in [31] for fraud detection within the Saman Bank in Iran; an e-banking security system employing Cryptography and Steganography was introduced in [25] for preventing online banking fraud.

Apart from the main implementations on transaction data, the authors in [28] focused on phishing detection from official banking websites and applied a multi-label classifier based associative classification DM for effective detection of phishing in websites with high levels of accuracy. In order to improve the customer credit card churning prediction for a Latin-American bank, the authors in [29] adopted improved DM techniques that are based on K-means clustering and support vector machines (SVM). Blog mining (text mining and cluster analysis) was applied in [30], where security risks, protection strategy and security trends of mobile banking were summarized from more than 200,000 results of the Google blog search engine.

There are also researchers who paid extra attention to money laundering detection. For instance, a DM model is presented in [6] that applied K-means clustering and Association Rule Mining for identifying suspected sequence of money laundering processes. A novel technique named Bitmap Index-based DT was proposed in [32] for evaluating the risk factor of money laundering with Statlog German credit data.

Table 1. Summary table of Data Mining applications in banking since 2013.

Sector	References	Key Techniques	Regions	Purposes	
Security and fraud detection	[6,10–32]	classification (DT, NN, SVM, NB), k-mean clustering, ARM	Australia [12], Latin-America [29], Greece [24], Germany [32], Belgium [21], UCI Repository [15–18]	Identifying phishing, fraud, money laundering, credit card fraud, security trend of mobile/online/traditional banking.	
Risk management and investment banking	[33–51]	classification (DT, NN, SVM, NB, LR), k-mean clustering	UCI Repository International Dataset [41,42,45], Australia [51], Iran [37], Indonesia [34], China [35], German [36,38,39,51], Taiwan [51], US [49], Canada [46]	Credit scoring, credit granting, risk management for peer-to-peer lending.	
Customer profiling and knowledge	[52,53]	classification (DT, NN), k-mean clustering	Jamaica [52]	Efficiently build accurate customer profiles.	
Customer segmentation	[54,55]	k-mean clustering	Iran [54]	Provide sufficient customer segmentation, conduct customer-centric business strategies.	
CRM	Customer satisfaction	[56]	k-mean clustering, classification (NN)	Spain [56]	Make the most strategic investment on maintaining and enhancing customer satisfaction.
Customer development and customization	[57–68]	classification (DT, NN, NB, LR, SVM), k-mean clustering	Portugal [57–62], Turkey [66,67], China [68], Taiwan [65], UCI Repository [64]	Strategic banking via direct marketing, targeted marketing, product cross/up selling.	
Customer retention and acquisition	[69–75]	classification (DT, NN, LR, SVM), ARM, k-mean clustering	EU [69], China [70], Nigeria [71], Croatia [73], Bangladesh [75]	Customer churn prediction and prevention, attracting potential customers and strategic future service design.	
Other advanced supports	[76–84]	classification (NN, DT, SVM), k-mean clustering	Nigeria [77], Turkey [78,81], Canada [80], ASEAN [82], Islamic banks [83], BRICS [84], US [79]	Branch strategy, bank efficiency evaluation, deposit pricing, early warning of failing bank.	

3.2. Risk Management and Investment Banking

Those interested in DM applications in credit scoring up until 2012 are referred to the review in [33].

A large proportion of recent applications applied DM techniques for credit scoring across the banking sectors globally [34–42]. A bank in Indonesia was investigated in [34]. Chen et al. [35] analyzed data from 16 listed Chinese commercial banks whilst the data set from the Export Development Bank of Iran was evaluated in [37]. Koh et al. [36] constructed a two-step method for credit scoring through the database in a German bank. Similarly, with the German credit scoring dataset, Harris [38] demonstrated the clustered SVM classifier for credit scoring, and Zhao et al. [39] presented the improved Multi-layer Perceptron NN model by employing the back propagation algorithm. The authors in [40] proposed the credit risk evaluation approach with the assistance of external evaluation and sliding window testing; it was verified on real life data from EDGAR. Later, Alaraj and Abbod [41,42] adopted the classifiers consensus system and proposed a hybrid model for credit scoring. The authors applied the classifier combination rule based on the consensus approach for experiments with seven real world credit data sets.

It is noteworthy that there are a few research studies that specifically targeted classification techniques and their applications in credit scoring leading to significant research developments. For instance, Lessmann et al. [43] reported on relevant research up until 2014 and conducted comprehensive experiments with real life Australian and German credit data sets for seeking the optimal classifier. Louzada et al. [44] recently produced a systematic review that specifically focused on the applications of classification techniques for credit scoring. Here, the main classification methods for credit scoring were summarized and introduced along with a detailed analysis of theoretical and paradigm trends. A recent comparative study in [45] conducted experiments on credit data sets from six different regions (including Australia, Germany, Iran, Japan, Poland and the US) to rank and identify the best classifier for credit scoring out of 25 different classifiers that were considered.

There are also researchers who focused on the decision making process of credit granting [46–48]. Specifically, a personal bankruptcy system was proposed in [46] for bad account prediction and was implemented on the credit card data set from a Canadian bank. The authors in [47] analyzed the capacity of credit union members at settling their commitments. A decision support system for banks was proposed in [48] for leading institutions to monitor account receivables and maintain profitability.

With regard to DM applications in risk management of peer-to-peer (P2P) lending, the authors in [49] focused on profit scoring by predicting the internal rate of return for a decision support system of P2P lending. The proposed model was verified by an experiment considering US Lending Club data. Similarly, the research in [50] adopted LR and K-means clustering techniques for detecting bad credit scores for P2P lending data. Recently, Xia et al. [51] employed three real life credit data sets and two P2P lending data sets for evaluating the performance of a newly proposed approach that employed extreme gradient boosting and Bayesian hyper-parameter optimization.

3.3. Customer Relationship Management (CRM)

CRM is “a comprehensive strategy and process of acquiring, retaining and partnering with selective customers to create superior value for the company and the customer” [85], and it has been overwhelmingly influenced by the DM techniques [86]. A previous review of DM applications in CRM with respect to a general research scale was published in [87], which thoroughly reviewed relevant literature up until 2008, and a recent general review can be found in [88]. However, with regard to the specific interest of the banking sector, only one short review can be found in [89] covering literature up until 2013.

The representative framework of customer analytics in banking is iCARE supported by IBM; more details of its solutions and a real case study for a commercial bank in Southeast China can be found in [90].

3.3.1. Customer Profiling and Knowledge

To build up an accurate customer profile, it is crucial for banks to extract valuable information from the customer behaviour with the assistance of DM techniques [91]. Mansingh et al. [52] demonstrated the use of DM techniques on survey data of internet banking users in Jamaica, which further aided in the decision making process of analyzing attitudinal, behavioral and demographic variables collectively for the purpose of prediction and profiling. By focusing on the data obtaining process, the online data imputation framework incorporating DM techniques was introduced in [53] and verified via an application of real banking data sets.

3.3.2. Customer Segmentation

In order to achieve better understanding of mobile banking customers and implementations of customer-centric strategies, Noori [54] proposed the customer segmentation model for an Iranian bank. Later, the authors in [55] introduced a framework based on the transaction driven parameters for proper segmentation of a bank's customers.

3.3.3. Customer Satisfaction

DM techniques have also been applied for maintaining and enhancing banking customer satisfaction, for instance, in Spain the core determinants of level of trust for banking customers are analyzed using DM techniques in [56].

3.3.4. Customer Development and Customization

A number of DM applications in customer development and customization focused on marketing related tasks. The bank direct marketing data sets from Portuguese banking institutions have been popular data resources and were investigated in [57–59] which compared the performances of four different DM classification techniques. The same dataset was used for verifying the proposed approach in [60] that employed correlation-based feature subsection selection algorithm along with the data set balancing technique. Later, this model was developed by [61] with an ensemble framework. Moreover, a profit driven artificial NN approach was proposed in [62] and a similar study applied two steps model of K-mean clustering and classification [63]. Recently, Lahmiri [64] proposed a two step system that combined a NN ensemble model and Particle Swarm Optimization for optimizing the initial weights of each NN in the ensemble framework. This was also verified by the bank direct marketing data with outstanding performance in relation to the baseline approaches.

Apart from the continuous exploration of the above Portuguese direct marketing dataset, Shih et al. [65] presented a target marketing model for commercial banks for the personal loan service, and the experiment was conducted with the data from a bank in Taiwan. With the direct marketing data set of a Turkish bank, Mitik et al. [66,67] proposed a two step hybrid system and achieved promising accuracy and a huge increase in the overall profit/cost ratio. Another regionally focused research by Wang and Petrounias [68] analyzed the relationships between demographic characteristics and mobile banking in China with big data collected through questionnaires. This application contributes to guiding the development of marketing strategies for domestic banks in China.

3.3.5. Customer Retention and Acquisition

DM techniques have been widely applied to detect early warnings in customer behaviour such as reduced transactions and account status dormancy, which will help the banks to provide proactive and strategic steps for preventing customer churn and switching banks.

Data relating to private banking customers in a European bank was analyzed in [69]; He et al. [70] employed the SVM technique for customer churn and attrition prediction with a real life data set from a Chinese commercial bank; customer records of a major bank in Nigeria were investigated in [71]; later, in [72] a NN classification technique was applied on the customer database of an international

bank for customer churn prediction; similar research was also conducted in a small Croatian bank in [73] and the electronic banking service data set in [74].

More recent research by Azad [75] investigated mobile banking adoption in Bangladesh by employing the NN technique on a structured questionnaire data set. The research identified the most influencing factor on mobile banking adoption, which then assisted in attracting potential customers and the design of future services.

3.4. Other Advanced Supports

Social media has been identified as another top trend in the banking sector [76], for instance, text mining was applied in [77] to extract hidden information from social media data in Nigeria (Twitter and Facebook) to assist the banking sector decision making and internet marketing. Specifically, the experiments analysed the unstructured data extracted from Facebook and Twitter for five of the largest Nigerian banks via Text Mining and K-means clustering.

A less mainstream application by Batmaz et al. [78] focused on the DM application for deposit pricing and identifying its main determinants. This research was conducted on the customer level data set of a commercial bank in Turkey, and beneficial conclusions for strategic deposit pricing were achieved, contrary to existing evidence obtained from macro-level bank data.

With regard to the evaluation of bank branch performance and providing early warning for failing banks, DM techniques like clustering and classification were adopted in [79–81], where [80] conducted the experiment on a real data set of a Canadian bank, while [81] on the Ziraat bank in Turkey. These studies assisted the regulatory bodies with early signals of banks/branches that require immediate attention; they also contributed to achieving strategic expansion design. Similarly, Wanke et al. [82–84] have conducted a series of research studies into banking efficiency evaluation by adopting NN techniques. Their applications have covered banking sectors from ASEAN, Islamic and BRICS countries.

4. Key DM Techniques, Software for Banking and Trends

Following the detailed review of over 100 publications by topics, the key DM techniques adopted in banking are identified, which include cluster analysis [92], association rule mining (ARM) [93] and classification techniques [94], which include but are not limited to Decision Trees (DT) [95,96], Neural Networks (NN) [97], Support Vector Machines (SVM) [98], Naive Bayes (NB) [99], and Logistic Regression (LR) [100]. Note that a brief introductory summary on these DM techniques can be found in [3,4].

This research also reveals the trends of DM applications and techniques in the context of banking based on the key information manually extracted from the identified recent applications. Note that the following statistics and diagrams are achieved based on the manually filtered information from the reviewed publications only and some applications did not clarify these specific information due to confidentiality related restrictions.

According to Figure 2, CRM applications account for about 35% of reviewed publications, which confirms the fact proposed by [101] that over 80% of financial service organizations globally list customer experience as its top priority. Within CRM applications, about half of the implementations are targeting customized marketing and cross/up selling, followed by customer retention and acquisition covering about 1/3 of CRM implementations. Recent research also addressed the significance of fraud detection and risk management as they account for 28% and 26% of the overall applications. This is due to the emerging need for combating cyber crime and developing more advanced technologies [102] (A review paper specifically targeting DM applications in fraud detection can be found in [103–105]).

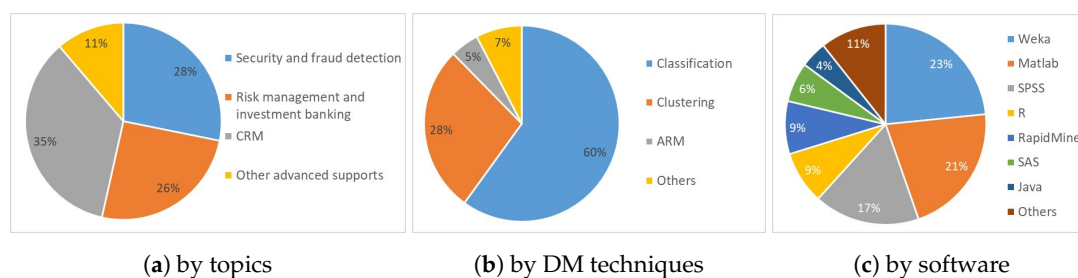


Figure 2. Key facts on DM applications in banking since 2013.

Weka, Matlab and SPSS are the most popular software adopted, followed by R and RapidMiner. Although there are about 30% of publications that did not clarify the software information, these facts are expected to assist the relevant research parties in finding corresponding analytical solutions or researchers who hold relative expertise. The most frequently adopted DM techniques are classification 60% and clustering 28%. However, it is noted that most of the applications are employing more than one DM technique, also generally the papers adopting classification techniques use more than one specific classification technique for comparison purposes. Specifically, K-mean clustering is the most frequently applied clustering technique and the top three classification techniques are NN, DT and SVM. ARM seems rarely exploited considering its 5% proportion, and it is hardly seen integrating text mining or social network analysis mining. Considering the availability of unstructured big banking data from customer profiles, feedback and call centre records [106], there exists a huge potentials for many DM techniques that have not been investigated before.

5. Conclusions

This paper successfully captured and systematically reviewed nearly 100 DM applications in banking post 2013. It fulfills the literature gap and serves as a quick reference guide for recent DM implementations in banking. Having reviewed these recent publications, it can be concluded that the banking sector has adopted DM mainly for fraud detection, risk management and CRM. In addition, most of the applications are using more than one DM technique, among which clustering and classification have shown sufficient evidence of both applicability and popularity.

Although the growing interest and promising performances have reflected the values and potentials of DM applications in banking, the obstacles of applying this techniques on big banking data is still noteworthy, for instance, the costly and time consuming process of personnel training towards pattern identification and data preprocessing, variable (feature) selection, complexity and difficulty of data quality assurance, large dataset storage and maintenance, etc.

Apart from the comprehensive summary of recent developments of DM applications in banking, this study also aims to present insights into the challenges and directions for future research. Firstly, it is noted that although the big banking data consists of large volumes of unstructured data, there are many DM techniques which continue to be rarely exploited, e.g., text mining, entity extraction, and social network analysis. This unbalanced exploration status can be caused by the limited access of big banking data, the shortage of researchers with relevant skill set, system constraints, and the lack of advanced data analytic tools [107]. Specifically, the confidentiality restrictions of banking related data have limited the progression of research. Therefore, seeking a proper solution for data availability will make a significant difference for future research.

In terms of the means of value creation by DM applications in banking, the banking sector has obtained sufficient rich customer information and the current implementations only focus on the marketing aspect. There exists a significant potential and valuable information waiting to be discovered. Moreover, a big proportion of available data channels like call center, customer surveys, and social media, are still waiting to be fully exploited. As a trending approach, machine learning methodology, especially deep learning, has been the emerging focus of a lot of scientific research.

Accordingly, it can be expected as another key direction for the banking sector in order to better embrace the era of big data.

Finally, as another suggestion of future work, the study of new analytical trends is also crucial for the banking sector, which aims to provide solutions for three types of use (i.e., Embedded Analytics, Citizen Data Science and Analytics for Data Scientists). Meanwhile, new technological trends in the era of big data can also continuously alter the research directions of DM applications in banking. For instance, the development of cloud computing can significantly improve the computational performance of most existing frameworks whilst the popularization of Internet of Things further enriches the big data resources, and it may also positively influence the embedded analytics and the development of dynamic big data analytics networks.

Author Contributions: All authors have contributed equally and substantially to the work reported.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mayer-Schonberger, V.; Cukier, K. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*; Houghton Mifflin Harcourt: New York, NY, USA, 2013.
2. Hassani, H.; Saporta, G.; Silva, E.S. Data Mining and Official Statistics: The past, the present and the future. *Big Data* **2014**, *2*, 34–43. [[CrossRef](#)] [[PubMed](#)]
3. Hassani, H.; Huang, X.; Silva, E.S.; Ghodsi, M. A review of data mining applications in crime. *Stat. Anal. Data Min. ASA Data Sci. J.* **2016**, *9*, 139–154. [[CrossRef](#)]
4. Hassani, H.; Huang, X.; Ghodsi, M. Big Data and Causality. *Ann. Data Sci.* **2018**, *5*, 133–156. [[CrossRef](#)]
5. Hassani, H.; Silva, E.S. Big Data: a big opportunity for the petroleum and petrochemical industry. *OPEC Energy Rev.* **2018**, *42*, 74–89. [[CrossRef](#)]
6. Kharote, M.; Kshirsagar, V.P. Data mining model for money laundering detection in financial domain. *Int. J. Comput. Appl.* **2014**, *85*, 61–64. [[CrossRef](#)]
7. Jayasree, V.; Balan, R.V.S. A review on data mining in banking sector. *Am. J. Appl. Sci.* **2013**, *10*, 1160. [[CrossRef](#)]
8. Pulakkazhy, S.; Balan, R.V.S. Data mining in banking and its applications—A review. *J. Comput. Sci.* **2013**, *9*, 1252–1259. [[CrossRef](#)]
9. Amani, F.A.; Fadlalla, A.M. Data mining applications in accounting: A review of the literature and organizing framework. *Int. J. Account. Inf. Syst.* **2017**, *24*, 32–58. [[CrossRef](#)]
10. Wongchinsri, P.; Kuratach, W. A survey-data mining frameworks in credit card processing. In Proceedings of the 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Mai, Thailand, 28 June–1 July 2016; pp. 1–6.
11. Bhasin, M.L. Menace of frauds in the Indian banking industry: An empirical study. *Aust. J. Bus. Manag. Res.* **2015**, *4*, 1–13. [[CrossRef](#)]
12. Wei, W.; Li, J.; Cao, L.; Ou, Y.; Chen, J. Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web* **2013**, *16*, 449–475. [[CrossRef](#)]
13. Akhilomen, J. Data mining application for cyber credit-card fraud detection system. In Proceedings of the Industrial Conference on Data Mining, New York, NY, USA, 16–21 July 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 218–228.
14. Carminati, M.; Caron, R.; Maggi, F.; Epifani, I.; Zanero, S. BankSealer: An online banking fraud analysis and decision support system. In Proceedings of the IFIP International Information Security Conference, Marrakech, Morocco, 2–4 June 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 380–394.
15. Malekpour, M.; Khademi, M.; Minae-Bidgoli, B. A Hybrid Data Mining Method for Intrusion and Fraud Detection in E-Banking Systems. *J. Comput. Intell. Electron. Syst.* **2014**, *3*, 1–6.
16. Seeja, K.R.; Zareapoor, M. FraudMiner: A novel credit card fraud detection model based on frequent itemset mining. *Sci. World J.* **2014**, *2014*. [[CrossRef](#)] [[PubMed](#)]

17. Zareapoor, M.; Shamsolmoali, P. Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia Comput. Sci.* **2015**, *48*, 679–685. [[CrossRef](#)]
18. Hegazy, M.; Madian, A.; Ragaie, M. Enhanced Fraud Miner: Credit Card Fraud Detection using Clustering Data Mining Techniques. *Egypt. Comput. Sci. J.* **2016**, *40*, 72–81.
19. Behera, T.K.; Panigrahi, S. Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural Network. In Proceedings of the 2015 Second International Conference on Advances in Computing and Communication Engineering (ICACCE), Dehradun, India, 1–2 May 2015; pp. 494–499.
20. Behera, T.K.; Panigrahi, S. Credit Card Fraud Detection Using a Neuro-Fuzzy Expert System. In *Computational Intelligence in Data Mining*; Springer: Singapore, 2017; pp. 835–843.
21. Van Vlasselaer, V.; Bravo, C.; Caelen, O.; Eliassi-Rad, T.; Akoglu, L.; Snoeck, M.; Baesens, B. APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decis. Support Syst.* **2015**, *75*, 38–48. [[CrossRef](#)]
22. Save, P.; Tiwarekar, P.; Jain, K.N.; Mahyavanshi, N. A Novel Idea for Credit Card Fraud Detection using Decision Tree. *Int. J. Comput. Appl.* **2017**, *161*, 6–9. [[CrossRef](#)]
23. John, S.N.; Anele, C.; Kennedy, O.O.; Olajide, F.; Kennedy, C.G. Realtime fraud detection in the banking sector using data mining techniques/algorithm. In Proceedings of the 2016 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 15–17 December 2016; pp. 1186–1191.
24. Azimi, A.; Noor Hosseini, M. The hybrid approach based on genetic algorithm and neural network to predict financial fraud in banks. *Int. J. Inf. Secur. Syst. Manag.* **2017**, *6*, 657–667.
25. Devadiga, N.; Kothari, H.; Jain, H.; Sankhe, S. E-Banking Security using Cryptography, Steganography and Data Mining. *Int. J. Comput. Appl.* **2017**, *164*, 26–30. [[CrossRef](#)]
26. Osinski, S.; Stefanowski, J.; Weiss, D. Lingo: Search results clustering algorithm based on singular value decomposition. In *Intelligent Information Processing and Web Mining*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 359–368.
27. Panigrahi, S.; Kundu, A.; Sural, S.; Majumdar, A.K. Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. *Inf. Fusion* **2009**, *10*, 354–363. [[CrossRef](#)]
28. Abdelhamid, N.; Ayesha, A.; Thabtah, F. Phishing detection based associative classification data mining. *Expert Syst. Appl.* **2014**, *41*, 5948–5959. [[CrossRef](#)]
29. Sundarkumar, G.G.; Ravi, V. A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Eng. Appl. Artif. Intell.* **2015**, *37*, 368–377. [[CrossRef](#)]
30. He, W.; Tian, X.; Shen, J. Examining Security Risks of Mobile Banking Applications through Blog Mining. In Proceedings of the 26th Modern Artificial Intelligence and Cognitive Science Conference (MAICS), Greensboro, CA, USA, 25–26 April 2015; pp. 103–108.
31. Hasheminejad, S.M.; Salimi, Z. FDiBC: A Novel Fraud Detection Method in Bank Club based on Sliding Time and Scores Window. *J. AI Data Min.* **2018**, *6*, 219–231
32. Jayasree, V.; Balan, R.S. Money laundering regulatory risk evaluation using Bitmap Index-based Decision Tree. *J. Assoc. Arab Univ. Basic Appl. Sci.* **2017**, *23*, 96–102. [[CrossRef](#)]
33. Siami, M.; Hajimohammadi, Z. Credit scoring in banks and financial institutions via data mining techniques: A literature review. *J. AI Data Min.* **2013**, *1*, 119–129.
34. Madyatmadja, E.D.; Aryuni, M. Comparative study of data mining model for credit card application scoring in bank. *J. Theor. Appl. Inf. Technol.* **2014**, *59*, 269–274.
35. Chen, Y.; Shi, Y.; Lee, C.F.; Li, M.; Liu, Y. Measuring and Predicting Systemic Risk in the Chinese Banking System. In Proceedings of the 2014 IEEE International Conference on Data Mining Workshop (ICDMW), Shenzhen, China, 14 December 2014; pp. 55–59.
36. Koh, H.C.; Tan, W.C.; Goh, C.P. A two-step method to construct credit scoring models with data mining techniques. *Int. J. Bus. Inf.* **2006**, *1*, 96–118.
37. Koutanaei, F.N.; Sajedi, H.; Khanabaei, M. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *J. Retail. Consum. Serv.* **2015**, *27*, 11–23. [[CrossRef](#)]
38. Harris, T. Credit scoring using the clustered support vector machine. *Expert Syst. Appl.* **2015**, *42*, 741–750. [[CrossRef](#)]
39. Zhao, Z.; Xu, S.; Kang, B.H.; Kabir, M.M.J.; Liu, Y.; Wasinger, R. Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Syst. Appl.* **2015**, *42*, 3508–3516. [[CrossRef](#)]

40. Danenas, P.; Garsva, G. Selection of support vector machines based classifiers for credit risk domain. *Expert Syst. Appl.* **2015**, *42*, 3194–3204. [[CrossRef](#)]
41. Alaraj, M.; Abbod, M.F. Classifiers consensus system approach for credit scoring. *Knowl.-Based Syst.* **2016**, *104*, 89–105. [[CrossRef](#)]
42. Alaraj, M.; Abbod, M.F. A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Syst. Appl.* **2016**, *64*, 36–55. [[CrossRef](#)]
43. Lessmann, S.; Baesens, B.; Seow, H.V.; Thomas, L.C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. Oper. Res.* **2015**, *247*, 124–136. [[CrossRef](#)]
44. Louzada, F.; Ara, A.; Fernandes, G.B. Classification methods applied to credit scoring: Systematic review and overall comparison. *Surv. Oper. Res. Manag. Sci.* **2016**, *21*, 117–134. [[CrossRef](#)]
45. Abellán, J.; Castellano, J.G. A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Syst. Appl.* **2017**, *73*, 1–10.
46. Xiong, T.; Wang, S.; Mayers, A.; Monga, E. Personal bankruptcy prediction by mining credit card data. *Expert Syst. Appl.* **2013**, *40*, 665–676. [[CrossRef](#)]
47. Sousa, M.D.M.; Figueiredo, R.S. Credit analysis using data mining: Application in the case of a credit union. *J. Inf. Syst. Technol. Manag.* **2014**, *11*, 379–396. [[CrossRef](#)]
48. Wu, D.D.; Olson, D.L.; Luo, C. A decision support approach for accounts receivable risk management. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *44*, 1624–1632. [[CrossRef](#)]
49. Serrano-Cinca, C.; Gutiérrez-Nieto, B. The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decis. Support Syst.* **2016**, *89*, 113–122. [[CrossRef](#)]
50. Li, H.; Zhang, Y.; Zhang, N.; Jia, H. Detecting the Abnormal Lenders from P2P Lending Data. *Procedia Comput. Sci.* **2016**, *91*, 357–361. [[CrossRef](#)]
51. Xia, Y.; Liu, C.; Li, Y.; Liu, N. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Syst. Appl.* **2017**, *78*, 225–241. [[CrossRef](#)]
52. Mansingh, G.; Rao, L.; Osei-Bryson, K.M.; Mills, A. Profiling internet banking users: A knowledge discovery in data mining process model based approach. *Inf. Syst. Front.* **2015**, *17*, 193–215. [[CrossRef](#)]
53. Nishanth, K.J.; Ravi, V. A computational intelligence based online data imputation method: An application for banking. *J. Inf. Process. Syst.* **2013**, *9*, 633–650. [[CrossRef](#)]
54. Noori, B. An Analysis of Mobile Banking User Behavior Using Customer Segmentation. *Int. J. Glob. Bus.* **2015**, *8*, 55.
55. Patel, Y.S.; Agrawal, D.; Josyula, L.S. The RFM-based ubiquitous framework for secure and efficient banking. In Proceedings of the 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), Noida, India, 3–5 February 2016; pp. 283–288.
56. Liebana-Cabanillas, F.; Nogueras, R.; Herrera, L.J.; Guillén, A. Analysing user trust in electronic banking using data mining methods. *Expert Syst. Appl.* **2013**, *40*, 5439–5447. [[CrossRef](#)]
57. Elsalamony, H.A. Bank direct marketing analysis of data mining techniques. *Int. J. Comput. Appl.* **2014**, *85*, 12–22. [[CrossRef](#)]
58. Moro, S.; Cortez, P.; Rita, P. A data-driven approach to predict the success of bank telemarketing. *Decis. Support Syst.* **2014**, *62*, 22–31. [[CrossRef](#)]
59. Bahari, T.F.; Elayidom, M.S. An efficient CRM-data mining framework for the prediction of customer behaviour. *Procedia Comput. Sci.* **2015**, *46*, 725–731. [[CrossRef](#)]
60. Vajiramedhin, C.; Suebsing, A. Feature selection with data balancing for prediction of bank telemarketing. *Appl. Math. Sci.* **2014**, *8*, 5667–5672. [[CrossRef](#)]
61. Amini, M.; Rezaeenour, J.; Hadavandi, E. A cluster-based data balancing ensemble classifier for response modeling in Bank Direct Marketing. *Int. J. Comput. Intell. Appl.* **2015**, *14*, 1550022. [[CrossRef](#)]
62. Zakaryazad, A.; Duman, E. A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing. *Neurocomputing* **2016**, *175*, 121–131. [[CrossRef](#)]
63. Barman, D.; Shaw, K.K.; Tudu, A.; Chowdhury, N. Classification of Bank Direct Marketing Data Using Subsets of Training Data. In *Information Systems Design and Intelligent Applications*; Springer: New Delhi, India, 2016; pp. 143–151.
64. Lahmiri, S. A two-step system for direct bank telemarketing outcome classification. *Intell. Syst. Account. Financ. Manag.* **2017**, *24*, 49–55. [[CrossRef](#)]

65. Shih, J.Y.; Chen, W.H.; Chang, Y.J. Developing target marketing models for personal loans. In Proceedings of the 2014 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Bandar Sunway, Malaysia, 9–14 December 2014; pp. 1347–1351.
66. Mitik, M.; Korkmaz, O.; Karagoz, P.; Toroslu, I.H.; Yucel, F. Data Mining Based Product Marketing Technique for Banking Products. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 12–15 December 2016; pp. 552–559.
67. Mitik, M.; Korkmaz, O.; Karagoz, P.; Toroslu, I.H.; Yucel, F. Data Mining Approach for Direct Marketing of Banking Products with Profit/Cost Analysis. *Rev. Soc. Strateg.* **2017**, *11*, 17–31. [[CrossRef](#)]
68. Wang, S.; Petrounias, I. Big Data Analysis on Demographic Characteristics of Chinese Mobile Banking Users. In Proceedings of the 2017 IEEE 19th Conference on Business Informatics (CBI), Thessaloniki, Greece, 24–27 July 2017; Volume 2, pp. 47–54.
69. Ali, O.G.; Ariturk, U. Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Syst. Appl.* **2014**, *41*, 7889–7903.
70. He, B.; Shi, Y.; Wan, Q.; Zhao, X. Prediction of customer attrition of commercial banks based on SVM model. *Procedia Comput. Sci.* **2014**, *31*, 423–430. [[CrossRef](#)]
71. Oyeniyi, A.O.; Adeyemo, A.B.; Oyeniyi, A.O.; Adeyemo, A.B. Customer churn analysis in banking sector using data mining techniques. *Afr. J. Comput. ICTs* **2015**, *8*, 165–174.
72. Ogwueleka, F.N.; Misra, S.; Colomo-Palacios, R.; Fernandez, L. Neural network and classification approach in identifying customer behavior in the banking sector: A case study of an international bank. *Hum. Factors Ergonom. Manuf. Serv. Ind.* **2015**, *25*, 28–42. [[CrossRef](#)]
73. Bilal Zoric, A. Predicting customer churn in banking industry using neural networks. *Interdiscip. Descr. Complex Syst.* **2016**, *14*, 116–124. [[CrossRef](#)]
74. Keramati, A.; Ghaneei, H.; Mirmohammadi, S.M. Developing a prediction model for customer churn from electronic banking services using data mining. *Financ. Innov.* **2016**, *2*, 10. [[CrossRef](#)]
75. Azad, M.A.K. Predicting mobile banking adoption in Bangladesh: A neural network approach. *Transnatl. Corp. Rev.* **2016**, *8*, 207–214. [[CrossRef](#)]
76. Suvarna, V.K.; Banerjee, B. *Social Banking: Leveraging Social Media to Enhance Customer Engagement*; Capgemini White Paper; Capgemini: Paris, France, 2014.
77. Afolabi, I.T.; Ezenwoke, A.A.; Ayo, C.K. Competitive analysis of social media data in the banking industry. *Int. J. Internet Market. Advert.* **2017**, *11*, 183–201. [[CrossRef](#)]
78. Batmaz, I.; Danisoglu, S.; Yazici, C.; Kartal-Koc, E. A data mining application to deposit pricing: Main determinants and prediction models. *Appl. Soft Comput.* **2017**, *60*, 808–819. [[CrossRef](#)]
79. Negnevitsky, M. Identification of failing banks using Clustering with self-organising neural networks. *Procedia Comput. Sci.* **2017**, *108*, 1327–1333. [[CrossRef](#)]
80. Herrera-Restrepo, O.; Triantis, K.; Seaver, W.L.; Paradi, J.C.; Zhu, H. Bank branch operational performance: A robust multivariate and clustering approach. *Expert Syst. Appl.* **2016**, *50*, 107–119. [[CrossRef](#)]
81. Met, I.; Tunali, G.; Erkoc, A.; Tanrikulu, S.; Dolgun, M.O. Branch Efficiency and Location Forecasting: Application of Ziraat Bank. *J. Appl. Financ. Bank.* **2017**, *7*, 1–13.
82. Wanke, P.; Kalam Azad, M.; Barros, C.P.; Hadi-Vencheh, A. Predicting performance in ASEAN banks: An integrated fuzzy MCDM–neural network approach. *Expert Syst.* **2016**, *33*, 213–229. [[CrossRef](#)]
83. Wanke, P.; Azad, M.A.K.; Barros, C.P.; Hassan, M.K. Predicting efficiency in Islamic banks: An integrated multicriteria decision making (MCDM) approach. *J. Int. Financ. Market. Inst. Money* **2016**, *45*, 126–141. [[CrossRef](#)]
84. Wanke, P.; Azad, A.K.; Emrouznejad, A. Efficiency in BRICS banking under data vagueness: A two-stage fuzzy approach. *Glob. Financ. J.* **2017**, *35*, 58–71. [[CrossRef](#)]
85. Parvatiyar, A.; Sheth, J.N. Customer relationship management: Emerging practice, process, and discipline. *J. Econ. Soc. Res.* **2001**, *3*, 1–34.
86. Berry, M.; Linoff, G. *Mastering Data Mining: The Art and Science of Customer Relationship Management*; John Wiley & Sons: New York, NY, USA, 1999.
87. Ngai, E.W.; Xiu, L.; Chau, D.C. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Syst. Appl.* **2009**, *36*, 2592–2602. [[CrossRef](#)]

88. Soltani, Z.; Navimipour, N.J. Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research. *Comput. Hum. Behav.* **2016**, *61*, 667–688. [[CrossRef](#)]
89. Desai, D.B.; Kulkarni, R.V. A Review: Application of data mining tools in CRM for selected banks. *Int. J. Comput. Sci. Inf. Technol.* **2013**, *4*, 199–201.
90. Sun, N.; Morris, J.G.; Xu, J.; Zhu, X.; Xie, M. iCARE: A framework for big data-based banking customer analytics. *IBM J. Res. Dev.* **2014**, *58*, 4:1–4:9. [[CrossRef](#)]
91. Gautam, P.; Singh, Y.P.; Shaikh, P. Significance and Importance of Data Mining for Marketing Analysis in Finance, Banking Sectors. *Int. J. Appl. Res. Sci. Eng.* **2017**, 26–29. Available online: <http://ijarse.org/images/scripts/201706.pdf> (accessed on 9 September 2017).
92. Xu, D.; Tian, Y. A comprehensive survey of clustering algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [[CrossRef](#)]
93. Agrawal, R.; Imieliński, T.; Swami, A. Mining Association Rules between sets of items in large databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data and ACM SIGMOD, Washington, DC, USA, 25–28 May 1993; pp. 207–216.
94. Pang-Ning, T.; Steinbach, M.; Kumar, V. *Introduction to Data Mining (2ed edition)*; Pearson: Boston, MA, USA, 2006.
95. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Belmont: Wadsworth, OH, USA, 1984.
96. Quinlan, J.R. *C4.5: Program for Machine Learning*; Morgan Kaufmann: Burlington, MA, USA, 1992.
97. Widrow, B.; Rumelhart, D.E.; Lehr, M.A. Neural networks: Applications in industry, business and science. *Commun. ACM* **1994**, *37*, 93–105. [[CrossRef](#)]
98. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
99. Langley, P.; Iba, W.; Thompson, K. An analysis of Bayesian classifiers. In Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992; Volume 90, pp. 223–228.
100. Dreiseitl, S.; Ohno-Machado, L. Logistic regression and artificial neural network classification models: A methodology review. *J. Biomed. Inform.* **2002**, *35*, 352–359. [[CrossRef](#)]
101. Marous, J. Improving the Customer Experience in Banking. Digital Banking Report. 2017. Available online: <https://www.digitalbankingreport.com/dbr/dbr246/> (accessed on 8 September 2017).
102. Lagazio, M.; Sherif, N.; Cushman, M. A multi-level approach to understanding the impact of cyber crime on the financial sector. *Comput. Secur.* **2014**, *45*, 58–74. [[CrossRef](#)]
103. Kirkos, E.; Spathis, C.; Manolopoulos, Y. Data mining techniques for the detection of fraudulent financial statements. *Expert Syst. Appl.* **2007**, *32*, 995–1003. [[CrossRef](#)]
104. Ngai, E.W.T.; Hu, Y.; Wong, Y.H.; Chen, Y.; Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decis. Support Syst.* **2011**, *50*, 559–569. [[CrossRef](#)]
105. Bhattacharyya, S.; Jha, S.; Tharakunnel, K.; Westland, J.C. Data mining for credit card fraud: A comparative study. *Decis. Support Syst.* **2011**, *50*, 602–613. [[CrossRef](#)]
106. Coumaros, J.; Buvat, J.; Auliard, O.; Roys, S.; Kvj, S.; Chretien, L.; Clerk, V. Big Data Alchemy: How can banks maximize the value of their customer data. In *Banks Have Not Fully Exploited the Potential of Customer Data*; Digital Transformation Research Institute and Capgemini Consulting: Paris, France, 2014.
107. Marous, J. Banking Industry Still Taking Small Steps with Big Data. The Financial Brand. 2017. Available online: <https://thefinancialbrand.com/64166/banking-big-data-advanced-analytics-ai/> (accessed on 3 September 2017).

