

Big Data Analytics for Improved Risk Management and Customer Segregation in Banking Applications

Subarna Shakya¹, S Smys²

¹Professor, Department of Electronics and Computer Engineering, Central Campus, Institute of Engineering, Tribhuvan University, Pulchowk, Lalitpur, Nepal

²Professor, Department of Computer Science and Engineering, RVS Technical Campus, Coimbatore, India

E-mail: ¹drss@ioe.edu.np, ²smys375@gmail.com

Abstract

While the phrase Big Data analytics is not only applicable for a certain realm of technology, diverse business segments like banking also benefit from the use of advanced mathematical and statistical models like predictive analysis, artificial intelligence, and data mining. If it is a query that is data volume generated in a bank or any financial institution is huge, it is absolutely a yes. As per the recent survey, it is observed that banks worldwide aren't just concentrating on improving the asset quality and fulfilling regulatory compliance but on the lookout for a digital convergence strategy to reach customers effectively in delivering services and products. As most of the data generated in internet banking and ATM transactions are unstructured accounting around for 2.5 quintillion bytes useful for fraud detection, risk management, and customer satisfaction, the use of trending Big Data Analytics methodology can be used to tackle the challenges and competition among banks. There are surplus advantages of Big Data strategy in the banking field and in this paper, we have made an analysis over Big Data Analytics on banking applications and their related concepts.

Keywords: Big Data, Big Data Analytics, Banks, Banking, Banking applications, Customer, Data Management

1. Introduction

Banking sector, is undoubtedly the backbone that supports companies worldwide [1]. Particularly, after the replacement of the barter system with monetary technique, banks play a vital role in various factors like key activities, investment & capital management, and project financing. Although most of the banks were nationalized in the interval 1960-1970, private banking became popular after 1991 due to globalization, and liberalization [2]. Among these, Citibank, Credit Suisse, JP Morgan, Wells Fargo, and UBS became the global leaders in the banking sector. Figure 1 is one such example of the banking structure followed in India. With the Reserve Bank of India as the primary head, 26 public, 43 foreign and 20 private banks are operating in India that is assumed to be the third-largest banking sector by 2025 [3].

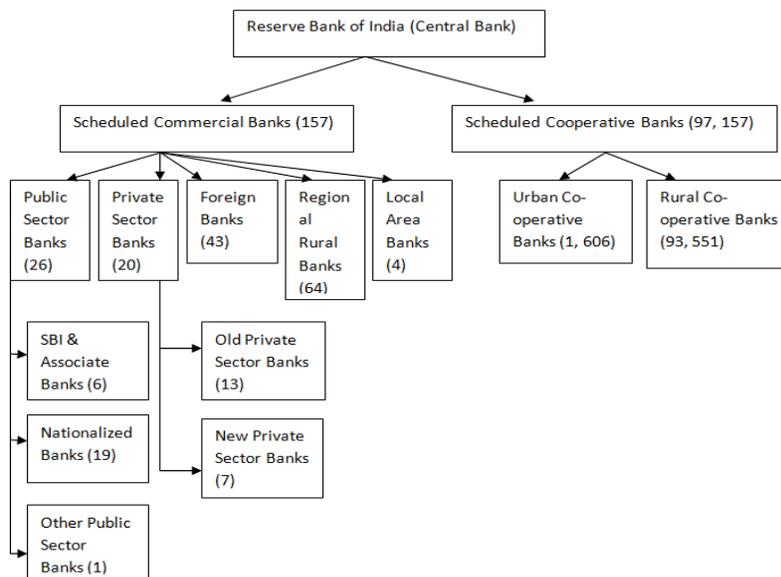


Figure 1. India's banking industry structure

Due to heavy competition between private and public banks, various innovative ideas and schemes were put forward to ease the dependence on physical branches and the reach of customers effectively [4]. Such visual and mobile-friendly strategies include remote, mobile, and internet banking where one great example for successful accomplishment of such non-branch channel transactions is HDFC bank [5]. With the successful accomplishment of 85% of transactions, the HDFC bank is recognized as a digital bank.

1.1 Big Data Analytics

As all the transactions and actions are data-driven, a huge number of data or millions of business records is generated in banks making it a difficult task for bankers and other people. To be clear, a high volume of data rushes from different transactions making risk management critical [6]. To overcome such issues, and tackle the data handling challenges, “Big Data” analytics is the right choice. Designed to fulfill the four factors- Veracity, Volume, Variability, and Velocity, Big Data is known for its phenomenal growth in surplus business and industry sectors. As each data gathered in the bank sector is to be utilized miserably, the “Big Data” method can be a game-changer in this process. Apart, it also plays a vital role in mitigating risks, improving extrapolative power, and promoting the growth of the BFSI sector, and financial institutions [7].

Although being a complicated or hectic process, with the help of Big Data Analytics, you can analyze or examine surplus different data sets and gather information like customer preferences, unknown correlations, and so on [8]. Such analytics also helps business experts in bringing out conclusions or decisions to move further. This full-fledged data-driven analytics approach is now being an integral part of the banking sector at present whereas the concept of “real-time evaluation” boosts profitability, and improves comprehensive customer understanding [9].

1.2 Profit of Big Data Analytics in Banking

From surplus advantages of Big Data Analytics on Banking, a few of them are listed here.

- **Complete analysis/acquiring of customer's expenditures & incomes:** With Big Data Analytics, you can access the clients or customer's expenditure or income thus by calculating factors like credit extensions, and risks [10].
- **Segregation of Customer base:** To suggest the right plans for clients/customers, segregating the total customer base into certain indicators is a vital part. With Big Data Analytics, you can make it effectively [11].
- **Fraud avoidance/Risk analysis:** Big Data has the potential to reduce the overhead costs implemented in the process of reporting, auditing, and verification by analyzing the stocks, transaction of the applicant [12].
- **Improve client faithfulness and feedback organization:** Big Data Analytics helps in effective management of feedbacks and increasing customer loyalty in BFSI sectors or financial institutions [13].

2. Banking- related Tools and Technologies of Big Data Analytics

As semi-structured and unstructured datasets couldn't be accommodated in traditional warehouses, we make use of Hadoop or other companion tools to transform them into NoSQL datasets. Such tools or technologies help in the easy organization of data types, collection, and other phrases of the analysis process. Popular types of tools and technologies associated are as follows:

1. **Kafka:** Kafka [14] belongs to the distributed subscribe/publish messaging system category. Designed to replace or to be considered as an alternative for the traditional message broker.
2. **YARN:** Recognized to be one of the key features in the second-generation Hadoop technology, YARN [15] is a cluster-management technology.
3. **HBase:** HBase [16] is a key/value data storage unit that is column-oriented to function on the HDFS.
4. **Hive:** Hive [17] is more popularly used in analyzing and querying huge datasets stored as Hadoop files making it the popular data warehouse infrastructure tool.
5. **Spark:** Spark [18] is based on the parallel-processing framework which can be used as an open-source option for high-scale datasets especially for the analytics applications in the clustered systems.

While the above-discussed ones are the popular tools for the Big Data Analytics technology in the banking sector, when it comes to real-time analytics in the Hadoop system, stream-processing engines like Flume, Apex, and Samza are in usage. Apart, there are also other cloud-based platforms like Map R, Amazon Web Services (AWS), and Microsoft Azure clouds that don't require any license [19].

When it comes to the technique used in Big Data Analytics in banking, data mining is the answer. Yes, consisting of 6 broad methods like classification, clustering, association, prediction, trees, and analysis, the brief description of each method is as follows:

1. **Classification:** Widely suited for fraud detection in the banking sector, classification is a common type of data mining technique in which the dataset is split or classified based on the preset examples [20].
2. **Clustering:** As its name suggests, it is the process of grouping the data based on criteria. In banking, it is the identification and grouping of similar data or transactions. For example, you can identify the clients or customers and group them under factors like service class (high-security policy demand), geographic location, service preferences, transaction method, and so on [21].
3. **Association Rule:** In association with various algorithms like DDA, CDA, and APRIORI, association rule helps ineffective management of the dataset. As the banking dataset is unstructured and mixed, there are times where transactions of the same account can occur twice or co-occur in the database [22]. In such circumstances, this rule can find the binary variables to uncover unwanted information.
4. **Prediction:** Prediction is a data mining method of Big Data analytics that is used to calculate or forecast the bond between independent and dependent variables [23]. For example, if money is an independent variable followed by a fraudster as a dependent, with the predictive analysis, you can detect any fraud attempt with the use of historical data.
5. **Decision/Random Trees:** To handle an array of problems of any category, decision and random trees are the influential data mining strategies [24]. With the motto to reduce overall data entropy, huge data is split into small ones in the decision tree, and the difference of possible errors & noise of the decision tree is carried in the random tree method.

6. **Link/Survival Analysis:** Both Link and Survival analysis are time-to-event analysis techniques that are mathematical-based models [25]. Constituting direct, and undirect mining, you can find valuable customers and predict hazard probabilities with the link or survival analysis.

3. Applications of Big Data Analytics in Banking

There are countless applications of Big Data Analytics in the banking sector or financial institutions and in this session, a few of them are discussed.

A. Fraud Discovery/Avoidance& Improved Risk Management

Banking, BFSI, or financial institution is prone to different hacks or breach and it isn't an unknown fact still now. Even though various efforts, methods, or strategies are in prevalence to get rid of them completely, the process is complicated and intricate. But, Big Data Analytics can put a full stop to it. Yes, in order to reduce the financial loss to an immense level, the above-discussed Big Data Analytics tools & technologies can help out that identifies massive datasets, analyze them, and detect the fraudster accounts or activity faster. The BI platforms can pinpoint the risks associated with the cash lending phase and deciding the interest rates.

Big Data is undoubtedly the best tool for the management of risks incurred in various metrics like operations, credit, commercial loans, and integrations. Offering accurate risk intelligence in real-time at a high rate, with the help of Big Data Analytics, the bank's risk organization team can get the below benefits:

- Build new opportunities or paths for the best price savings.
- Advance the value of scheme response time.

- Provide wide risk coverage
- Augment the parameters of risk models exponentially.

Such risk management metric is calculated using the tracking and reporting factors and they are

1. MIS/Regulatory Reporting
2. Disclosure Reporting
3. Real-time Keyboard Conversation Tracking
4. Anti-money Laundering

B. Improved Employee Engagement and Customer Segregation

While the bank growth metrics are not only related to the customer's strength, it also depends on the bank's employees. Whereas, Big Data not just offers technical credits by providing fraud and risk management, it also helps in analyzing the employee experience of the bank. With the help of Big Data tools, one can track and monitor the performance and involvement of each employee to identify the active performers and unhappy employees. Such application of Big Data not only increases productivity by reducing the workload but can upgrade to a better work culture as well.

As a result of the digitalization of financial services, customer insight and interaction with the financial institution or BFSI is increasing at present. By implementing Big Data Analytics, one can make use of a digital platform like social media, sensor technology, mobile devices, and scientific instruments to build interaction or promotion with the customers related to products or services. Along with the creation of an in-depth understanding of customer preferences, needs, and

customer experience, Big Data can keep you stay updated and on top of the competitive market. Unlike risk management, certain customer-centric factors contribute to customer management as well and they are mentioned in Table 1.

Table 1. Customer-Centric or Management Metrics

Customer life event analysis	Next best offer
Customer gamification	Client experience closed-feedback loop
Real-time allocation based offerings	Micro-segmentation
Quality of lead analysis	Sentiment analysis-enabled strategy management
Sentiment analysis-enabled lead/referral management	Sentiment-analysis-enable sales forecasting

C. Enhanced Market Trading Analysis and Personalized Banking Solutions

As market trading and transaction in the bank sector takes a long time and is a slow process, it can be fastened or the demand can be speedier when Big Data technology is put into practice. Although the strategies are a little complicated, with the use of Big Data tools, data gathered from different categories like asset classes, hybrid datasets, and market types can be analyzed in a fraction of second. Apart, aspects like risk factors, trade execution, profit & loss, and trends can be covered in a 360^o view.

4. Results and Discussion

As per the recent research and survey carried out with the banks and the customer experience, certain interferences are made that are discussed in this section. When Big Data



Analytics was put into practice, it is proved that the customer insight was meaning to 77% and it's a No to 23%.

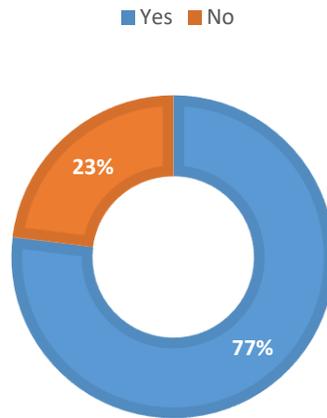


Figure 2. Customer Insights

Also, the results incurred based on the potential benefits driven from Big Data Analysis account to 5 factors like better customer management, increase in profit, gaining competitive edge, increase in business volume, and others.

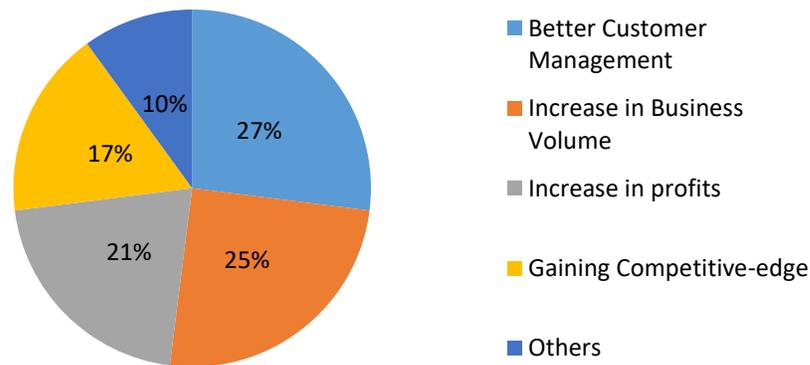


Figure 3. Benefits of Big Data Analytics

Not just that, further recommendation of the better utilization of Big Data Analytics in banking also reveals the following factors to be implemented in the upcoming days. Among which comes – creating robust data warehouses, educating the customers, training the employees, leveraging cloud, and using incisive data analysis engines.

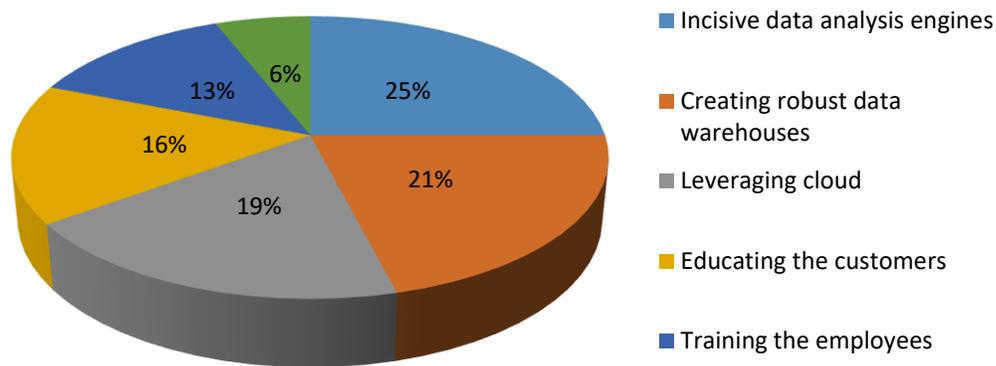


Figure 4. Recommendations for future Big Data Analytics

5. Conclusion

Years ago, it was a traditional type of banking process where an executive greets the customer or client on respective channel branches and fulfills their banking needs. But, at present, everything has changed upside down as client's needs, preferences, and geographic location varies time-to-time. In such a situation, Big Data Analytics is an all-in-one tool to get insight regarding factors like needs, lifestyle, fraud risks, and remedies. In this paper, a deep analysis of the Big Data Analytics on the banking sector, tools, technologies, and applications was discussed. Also, the expected result or practical interference to the expected result is given as well. Although Big Data Analytics is more important in fraud, and risk management, especially can yield the desired outcome when executed real-time with AI, still, the majority of the banks in India haven't started

their attempt towards such Big Data Analytics. With further development to connect different departments in a bank through the Big Data infrastructure, the research still goes on.

References

- [1] Manish, P. M., Kasale, S., & Simon, A. D. (2018). Banking & big data analytics. IOSR Journal of Business and Management, 55-58.
- [2] Smys, S., and Jennifer S. Raj. "Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 24-39.
- [3] Gupta, Sameer, Deepa Gupta, and Ruchika Bathla. "A Security Model for Enhancement of Social Engineering Process with Implementation of Multifactor Authentication." In International conference on Computer Networks, Big data and IoT, pp. 979-987. Springer, Cham, 2019.
- [4] More, R., & Moily, Y. Big Data Analysis in Banking Sector. International Journal of Engineering Research and Applications.
- [5] Joe, Mr C. Vijesh, and Jennifer S. Raj. "Location-based Orientation Context Dependent Recommender System for Users." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 14-23.
- [6] Srivastava, U., and Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian banks. Procedia Computer Science, 50, 643-652.
- [7] Hamdan, Yasir Babiker. "Faultless Decision Making for False Information in Online: A Systematic Approach." Journal of Soft Computing Paradigm (JSCP) 2, no. 04 (2020): 226-235
- [8] Zhu, X., & Yang, Y. (2021). Big Data Analytics for Improving Financial Performance and Sustainability. Journal of Systems Science and Information, 9(2), 175-191.

- [9] Saxena, Komal, and Arshdeep Singh Pahwa. "Enhanced Financial Module in Intelligent ERP Using Cryptocurrency." In International conference on Computer Networks, Big data and IoT, pp. 959-967. Springer, Cham, 2019.
- [10] Haoxiang, Wang, and S. Smys. "Big Data Analysis and Perturbation using Data Mining Algorithm." *Journal of Soft Computing Paradigm (JSCP)* 3, no. 01 (2021): 19-28.
- [11] ALI, Q., SALMAN, A., YAACOB, H., ZAINI, Z., & ABDULLAH, R. (2020). Does big data analytics enhance sustainability and financial performance? The case of ASEAN banks. *The Journal of Asian Finance, Economics, and Business*, 7(7), 1-13.
- [12] Thilaka, B., Janaki Sivasankaran, and S. Udayabaskaran. "Optimal Time for Withdrawal of Voluntary Retirement Scheme with a Probability of Acceptance of Retirement Request." *Journal of Information Technology* 2, no. 04 (2020): 201-206.
- [13] Jayaraj, T., and J. Abdul Samath. "Cloud Based Heterogeneous Big Data Integration and Data Analysis for Business Intelligence." In International conference on Computer Networks, Big data and IoT, pp. 926-933. Springer, Cham, 2019.
- [14] Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, 42, 13-24.
- [15] Shakya, Subarna. "A Self Monitoring and Analyzing System for Solar Power Station using IoT and Data Mining Algorithms." *Journal of Soft Computing Paradigm* 3, no. 2: 96-109.
- [16] Garg, Karan, and Rejo Mathew. "Review of Software Defined Networking Based Firewall Issues and Solutions." In International conference on Computer Networks, Big data and IoT, pp. 909-916. Springer, Cham, 2019.
- [17] Sivaganesan, D. "A Data Driven Trust Mechanism Based on Blockchain in IoT Sensor Networks for Detection and Mitigation of Attacks." *Journal of trends in Computer Science and Smart technology (TCSST)* 3, no. 01 (2021): 59-69.

- [18] Marr, B. (2017). *Data strategy: How to profit from a world of big data, analytics and the internet of things*. Kogan Page Publishers.
- [19] Smys, S., and Haoxiang Wang. "Security Enhancement in Smart Vehicle Using Blockchain-based Architectural Framework." *Journal of Artificial Intelligence* 3, no. 02 (2021): 90-100.
- [20] Rahman, M. R. (2021). *Algorithmic Trading Using Trend Following Strategy: Evidence from Indian Information Technology Stocks*. *Data Science and Data Analytics: Opportunities and Challenges*, 293.
- [21] Prasanna, I. Philo, and M. Suguna. "Detection of Distributed Denial of Service Attack Using NSL-KDD Dataset-A Survey." In *International conference on Computer Networks, Big data and IoT*, pp. 866-875. Springer, Cham, 2019.
- [22] Smitha, T. V. "A Study on Various Mesh Generation Techniques used for Engineering Applications." *Journal of Innovative Image Processing* 3, no. 2: 75-84.
- [23] Sfaxi, L., & Aissa, M. M. B. (2021). *Designing and implementing a Big Data benchmark in a financial context: application to a cash management use case*. *Computing*, 1-23.
- [24] Suma, V. "Community Based Network Reconstruction for an Evolutionary Algorithm Framework." *Journal of Artificial Intelligence* 3, no. 01 (2021): 53-61.
- [25] Sundarakumar, M. R., Mahadevan, G., Somula, R., Sennan, S., & Rawal, B. S. (2021). *An Approach in Big Data Analytics to Improve the Velocity of Unstructured Data Using MapReduce*. *International Journal of System Dynamics Applications (IJSDA)*, 10(4), 1-25.

Author's biography

Subarna Shakya is currently a Professor of Computer Engineering, Department of Electronics and Computer Engineering, Central Campus, Institute of Engineering, Pulchowk, Tribhuvan University, Coordinator (IOE) , LEADER Project (Links in Europe and Asia for engineering, education, Enterprise and Research exchanges), ERASMUS MUNDUS. He received MSc and

PhD degrees in Computer Engineering from the Lviv Polytechnic National University, Ukraine, 1996 and 2000 respectively. His research area includes E-Government system, Computer Systems & Simulation, Distributed & Cloud computing, Software Engineering & Information System, Computer Architecture, Information Security for E-Government, Multimedia system.

S Smys received his M.E. and Ph.D. degrees in Wireless Communication and Networking from Anna University and Karunya University, India. His main area of research activity is localization and routing architecture in wireless networks. He serves as Associate Editor of Computers and Electrical Engineering (C&EE) Journal, Elsevier, and Guest Editor of MONET Journal, Springer. He served as Reviewer for IET, Springer, Inderscience and Elsevier journals. He has published many research articles in refereed journals and IEEE conferences. He has been General chair, Session Chair, TPC Chair and Panelist in several conferences. He is Member of IEEE and Senior Member of IACSIT wireless research group. He has been serving as Organizing Chair and Program Chair of several International conferences and in the Program Committees of several International conferences. Currently, he is working as Professor in the Department of Information Technology at RVS technical Campus, Coimbatore, India.