

Using Big Data Analytics to Develop Marketing Intelligence Systems for Commercial Banks in Egypt

Sara Mohamed Asaad El Banna¹ and Prof. Nevine Makram Labib¹

¹Computer and Information Systems Department, Sadat Academy for Management Sciences, Maadi, Cairo, Egypt

Abstract. Nowadays, Big Data (BD) Analytics is receiving great attention in banking industry, considering the worthy data that have been stored for several decades, to reach the main targets of marketing by increasing the bank's efficiency of studying their clients, knowing their feedback, in addition to promoting active and passive security systems. This study focuses on utilizing BD analytics to develop marketing intelligence systems. It aims to explore the big data as a valuable resource for Egyptian commercial banks, to improve the customer experience, customer segmentation and profiling, selling products based on profiling, and describing customer behavior. In order to develop the proposed system, data were collected from several banks of transaction performed in 2016, including a report on customer satisfaction, a procedure of analyzing customer satisfaction data, consisting of about 39,000 records of transactions for customers and a collection of about 4,000 records of transaction data for cardholders. These data were analyzed using Apache Hadoop to perform many tasks such as profiling the bank's clients to groups, customer segmentation based on client's history, interest and habits, predicting customer behavior based on profiling, designing a new marketing strategy, and presenting the right offers to the bank's clients as individuals or as groups. It was concluded that BD analytics were very beneficial for achieving Marketing Intelligence in Banks.

1 Introduction

Nowadays Big data (BD) has got an extraordinary attention from the academic world and the IT business. In the automated world, data are created and collected with high speed. At present, more than 2 billion people worldwide are connected to the Web yet more than 5 billion people own cell phones. By 2020, 50 billion gadgets are anticipated to stand associated with the Web. by then, expected data production will be 44 times bigger than in 2009[1]. As data are exchanged and shared at light speed on optical fiber and wireless networks, both volume of information and the speed concerning market development increase. However, the expedient increment rate with respect to such large data produces various difficulties, such as the rapid growth of data, exchange speed, differing data, and security.

1.1 Big Data Definition

Big Data (BD) may be defined as "the amount of dynamic, large and scattered information generated by individuals, tools and machines"[2]. It needs innovative and extensible technology to collect, host and analyze a large amount of gathered information to capture insights

into real-time practices that relate to consumers, risk, profit, achievement, production, and management. It can also be identified as "a collection of information from both traditional and digital sources within and outside the association that is a source of discovery and continuous analysis"[3].

As for the characteristics of BD, certain elements are common across the different definitions, namely the 5Vs Volume, Velocity, Variety, Veracity and Value.

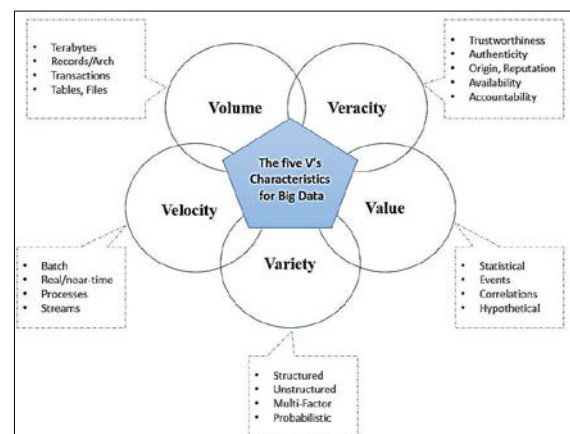


Fig. 1. Characteristics of BD -5Vs [4]

* Corresponding author: saramelbanna@gmail.com

Volume refers to the great preparation of information that is generated every second. This makes growing information sets huge for storage and analysis using traditional database technology for **Velocity**, it refers to the speed with which modern information is established and the speed at which information moves as for **Variety**, it refers to non-similar forms of information that we can use at this moment. With BD technology, we can now benefit from similar forms of information, including conversations, social networks, images, sensor data and audio and combined with more organized traditional structured data. concerning **Veracity**, it refers to confusion or reliability of information. With a lot of great information, quality, and accuracy, you can control. **Value** refers to the ability to change the data to an amount. It is important that the institutions make a claim to any endeavor to collect and benefit from the great information.[5].

1.2 Applications of BD in Banking Sector

Following are some of the recent application Of BD In Banking Sector.

Fraud Detection: Banks use analytics to detect patterns and determine their type if they are normal or abnormal. By analyzing large data, banks will have a fraud detection tool that will detect patterns, which will raise an early alarm making the analysis system suggest taking immediate actions that in turn will stop fraud before it happens.

Real-time Risk Analysis: A real-time analysis of transactions will support banks against capital, credit risk and financial risks when they skip their initial beginnings and alert banks to take suitable action. BD could be an appropriate tool because of its high frequency in terms of business data and sheer volume.

Client Satisfaction and Earning Loyalty : BD's ability to make the customer's DNA through the history of dealing, the rate of debt, and repayment habits, all tell a story and can facilitate the bank to distinguish customers which makes us understand the behavior of customers. Through this understanding, we can classify customers and enable large data to aggregate different clients into segments will create based on customer behavior.

Personalized Marketing: Customer segmentation will be more utilized produce and create new schemes, aimed immediately to the particular needs of their customers. By learning behaviors return from BD sources and analyzing past and current cost and transactions, a bank will get a clear understanding of the way to get the best response frequency from their customers. The access to information of selection associated price creates a chance of providing customized goods to customers can cater to an This will raise satisfaction levels for customers , will improve profit for banks and this will offers banks the flexibility to make additional relationships with clients.

Create New Product: By learning the direction of client needs and satisfaction through analytics you can create products according to the wants of clients.

Reduce costs: BD is best seen as a price cutting live within the banking sector. It's data management system that may scale back the necessity for storage making easy for data and lower the computational needs for the bank. The convoluted and learning operations of massive information technologies will be accessed and managed by average ability employees to avoid hiring specialist operators. The utilization of machine learning suggests that the system once set the suitable parameters, will learn for itself to relate and analyze information to create it additional helpful to the bank.

Increase Bank Efficiency: The use of resources to attain outcome are often managed and even reduced thanks to BD. the major objective is to scale back the time employed in accessing, decoding and manipulating data and BD would cut back hiding and increase turnout. The banking systems would be economical and secure with abundant less time wasted throughout background checks, lines in banking halls, extended application procedure and redundant authentication transaction[6].

1.3 Problem Definition

The main problem in this research lies in the fact that Commercial banks are still using traditional methods and do not make use of BD.

Exploratory studies showed that banks in Egypt are not using new technologies to develop marketing strategies. This makes them lose many competitive advantages in the marketing area. As a result, banks need to shift the paradigm for designing, developing, deploying, and maintaining BD solutions supported by a wave of emerging technologies that provide the flexibility and scalability required to support the shift.

1.4 Objective of The Study

The main objective of the study is to apply BD analytics on the valuable resources for Egyptian commercial banks in order to develop Marketing Intelligence Systems in banks based on customer profiling.

This should eventually lead to improving customer service and adding new products and services based on customer profiling.

1.5 Research Methodology

The study involves interviews with 5 CIOs and 5 IT managers from Egyptian commercial banks to find out if computer technologies affect marketing decisions in banking and how they affect marketing strategies of the Bank.

It also makes use of a survey for customers to find out their satisfaction levels with the products' offers of the bank and their payment preferences (cash – credit card). In addition, data we extracted from customers transactions. All of the previously - mentioned data are analyzed.

2 Literature Review

Following are some of the most recent studies that dealt with BD in banking sector.

A study explored ways on how to improve the efficiency of the banking system with BD. It covered customer satisfaction analysis, by identify the main customer of the bank with the analysis link, improve customer feedback, detected when the customer is about to leave and improve security enhancement. It found out that if the banking industry can effectively implement all analyzes and algorithms (feeling, aggregation, correlation, survival, decision tree, analysis and data file), it will increase the efficiency of the banking system[7].

Another study in the financial services industry dealt with a particular focus on the banking industry, highlighting many of the business problems that have been resolved through BD analysis. It also covered some of the latest technologies, which combined with large data analytics, have significantly changed the quality and productivity of the banking industry. The disciplines of the stock market and insurance have been explored, with predictive and directional analyzes playing a key role. The paper concluded with some challenges, which hamper the full implementation of BD analyzes in the financial services industry to enhance business value[8].

Third study made use of BD effectively to gain tremendous value in customer retention, targeted marketing, fraud reduction, internal operation, and many other areas. Customers' data collected across diverse systems enable advanced sectors to drive new customer growth and improve marketing ROI. Excessive analytics facilitate the opportunities for mutual forecasting and growing sales to increase portfolio share and long-term customer value. It was found out that BD also plays an important role in improving operational efficiency. Inter-institutional data supports regulatory requirements, new requirements and early detection of fraud. [9].

A fourth study explained that BD analysis is being implemented across many areas of Banking sector, and helps them to provide higher services to their clients, all Internal and external, besides improving them. This study explains the impact and advantages of BD analytics on banking sector through analyzing transactional and sentiment analysis for the Banking Sector. There are other ways that banks and various monetary institutions can obtain the knowledge associated with the client in order to analyze emotions, ranging from social media sites for various marketing research channels[10].

A final study showed the relationship between Customer Relationship Management (CRM) and BD. Although CRM deals with understanding human behavior and interests, BD can be expected to improve customer relationships as it allows for multi-directional communication and customization. BD enhances CRM strategies by better understanding customer habits and behaviors so that business can offer CRM to be more personalized and customized for each customer. Finally, CRM will make huge data with tools and better strategies more personalized to customers because they

understand the well targeted audiences and the intended messages to be sent. The previous studies showed that BD is important to implement in banks, especially in developing marketing strategies[11].

3 The Proposed Marketing Intelligence System

Following are the stages of Marketing Intelligence System implementation.

3.1 Data Collection

3.1.1 Data Collection Methods

1. Interviews with 5 CIOs and 5 IT managers of Egyptian commercial
2. Survey for customers to see their satisfaction with the products offers by the bank and the way they prefer to buy products (cash – credit card).
3. Access to transactions for the customers.

3.1.2 Data sources

1. Customer satisfaction Report

It was found that 20% of customers still don't know the products that can be purchased through the bank, 37% are not satisfied from the services provided by the bank through credit cards or customer service. 43% of customers have satisfaction with the services of the bank.

2. Tables

The data were collected from the bank for clients in 2016 the data are in CSV format dataset about (39,000 record) transactions for and (4,000 record) for cardholders.

These data were analyzed by using Apache Hadoop because it is scalable, cost effective, flexible, fast, and resilient to failure[14].

3.2 Tool Used :Apache Hadoop

It is collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of BD using the MapReduce programming model. It is originally designed for computer clusters built from commodity hardware still the common use it has also found use on clusters of higher-end hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common occurrences and should be automatically handled by the framework. The base Apache Hadoop framework is composed of the following modules:

Hadoop Common : contains libraries and utilities needed by other Hadoop modules.

Hadoop Distributed File System (HDFS) : a distributed file-system that stores data on commodity machines, provides high-throughput access to application data.
 Hadoop YARN : is a platform responsible for job scheduling and cluster resource management. Hadoop MapReduce an implementation of the MapReduce programming model for parallel processing of large data sets[12].

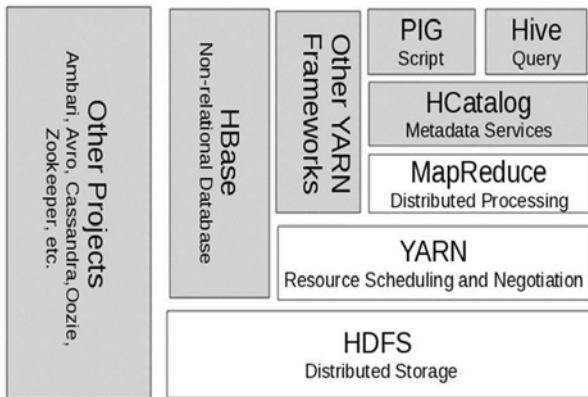


Fig. 2. Apache Hadoop Ecosystem [13].

The following figure illustrates the steps of BD Analytics

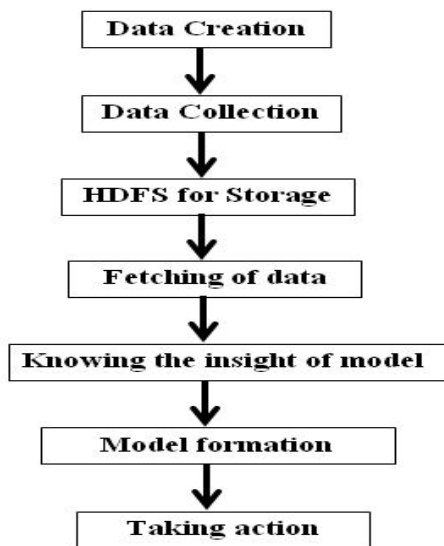


Fig.3. Steps of BD Analytics in Banking

3.2.1Data Creation: Banks provide new services on a continuous basis and thus daily increase in the number of customers at banks and all its banking establishments are working on the creation and collection of a large amount of data and transactions can be used in BD analytics.

3.2.2Data collection: The data were collected in 2016 from an Egyptian bank including a report on customer satisfaction, a procedure of analyzing customer satisfaction data, consisting of around 39,000 records of

transactions for customers and a collection of about 4,000 records of transaction data for cardholders.

3.2.3HDFS for Storage: Distributed file system that stores data on commodity machines provides high-throughput access to application data. One of the components of apache hadoop as for the application of the model, we used one of the IBM products (IBM BigInsights Virtual Machine) for BD analytics.

3.2.4Fetching of data: Uploading the data to HDFS.

3.2.5Knowing the insight of model: By creating Java Map/Reduce Program.

3.2.6Model formation: Using java coed to categorize a client into three categories high, medium and low.

3.2.7Tacking action: After analysis the data we can give the decision-maker all the result to development Marketing Intelligence.

4 Results and Discussion

The following transaction types have been taken into consideration: Allowance / salary – Account transfer - ATM – Cared card (Debit –Credit) contains (4,000 records) of transaction data for cardholders – Balance From the following analysis, the following results were concluded:

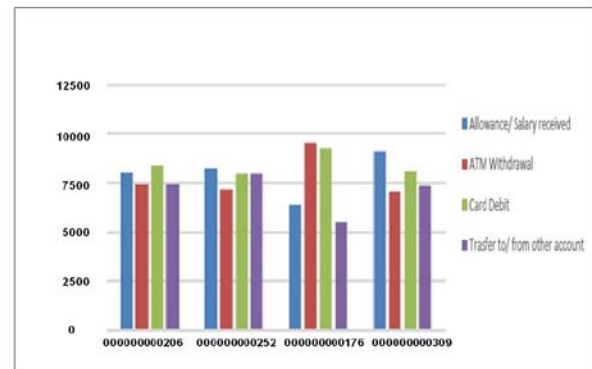


Fig. 4. Transaction Types

4.1.1 Client account_ID ending with 206 and 252 are for employees who spend money as soon as they receive it.

4.1.2 This person can plan for savings later or be given a return from the contract.

4.1.3 The number of the client account_ID that ends with the number 176 is expected to be paid monthly as it is transferred to his account every first month and this coincides with the salary dates in companies.

4.1.4 This person can be sold an investment plan, since his income is regular, and depending upon his

debit frequency, can be sold a credit card offering high credit capabilities.

4.1.5 Transaction data may also be used to estimate consumers who can sell types of financial products. It is frequently used by banks to divide and target potential customers.

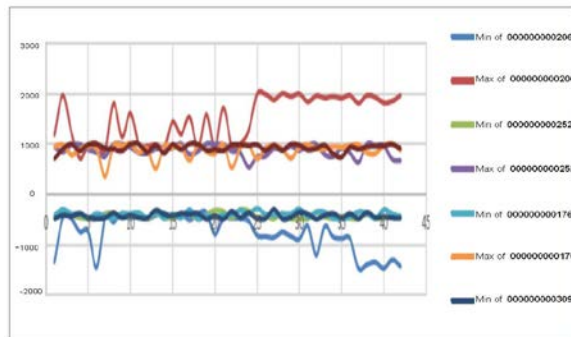


Fig. 5. Customer Behavior Pattern

4.1.6 The client account_ID which ends with a number 206 connected to the upper limit of spending based on consumer behavior analysis.

4.1.7 This person has a capacity to spend, and though infrequently, he/she has spent more than what is credited to his account during certain periods.

4.1.8 The person is also a perfect candidate to use a credit card. A credit card can be provided or approved. If it already uses one, its credit limit can be increased. Offers associated with a credit card can be extended to this person as they are likely to use their cards.

5 Conclusion of The Study

Based on using BD to understand customer behavior, interests and as well as their spending patterns, the following was performed:

1. Profiling the bank's clients to groups such as (homogeneous or heterogeneous)
2. Customer segmentation based on client (history, interest and habits).
3. Predicting customer behavior based on profiling
4. Designing a new marketing strategy
5. Presenting the right offers to the bank's clients as individuals or as groups

This resulted in developing Marketing Intelligence Systems for Bank.

6 Future work

Following are some research directions:

1. Using BD analytics across different sectors of the banking industry.
2. Develop credit risk models that enable consumer and commercial collections practices to prioritize their

actions according to propensity for delinquency or payment.

3. Using analytics in Fraud Detection systems so as to detect patterns and types.
4. Applying real-time analysis.
5. Including both semi-structured and unstructured data in the analysis.

References

1. N. Khan et al., *Big Data : Survey , Technologies , Opportunities , and Challenges*, vol. **2014**, (2014).
2. EY, *Big data Changing the way businesses*, Int. J. Simul. Syst. Sci. Technol., vol. **16**, no. April, p. 28, (2014).
3. L. Aurher, *What is Big Data?*, (2013), (on-line) <https://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/#6df9b6155c85>
4. S. Vanitha and P. Balamurugan, *BIG DATA AND HADOOP TECHNOLOGY-A STUDY*, vol. **8**, no. 554, pp. 554–561, (2018).
5. G. Bailey, *5 Vs for big data in project management*, (2018), (online): <http://www.prosymmetry.com/big-data-in-project-management/>
6. M., *The role of big data in the banking industry*, (2016), (on-line): <https://bigdata-madesimple.com/role-big-data-banking-industry/>
7. S. H. Oluwasola, *How to Improve Efficiency of Banking System with Big Data (A Case Study of Nigeria Banks)*, vol. **6**, no. 6, pp. 2015–2018, (2017).
8. V. Ravi and S. Kamaruddin, *Big Data Analytics Enabled Smart Financial Services: Opportunities and Challenges*, 5th Int. Conf. BDA 2017, Hyderabad, India, vol. **1**, pp. 15–39, (2017).
9. S. Hk, *Big Data & Analytics : Tackling Business Challenges in Banking Industry*, vol. **8**, no. 2, (2017).
10. I. Journal, *Impact of Big Data analytics on banking sector*, vol. **5**, no. 11, pp. 3138–3141, (2016).
11. M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh, *Customer relationship management and big data enabled: Personalization & customization of services*, Appl. Comput. Informatics, no. August, (2018).
12. Apache Software Foundation, *Apache Hadoop*, (2018), (on-line): <https://hadoop.apache.org/old/>
13. D. Eadline, *What is Apache Hadoop Yaran and Why Shoud You Cara*, (2015), (on-line) <http://www.informit.com/articles/article.aspx?p=2303783>
14. R. Prasad, *What is the advantages of Hadoop and Big data?* (2017), (on-line): <https://www.mindsmapped.com/hadoop-advantages-and-disadvantages/>