

Adithya Venkatramanan

Dr. Anita Wholuba

Composition II

1 April 2023

The Implications of Data Mining in Banking

As global citizens in the 21st century, we are living in a world surrounded by information. Big data analytics is the process of analyzing large datasets to find trends and draw meaningful conclusions (Smaya 2). According to Clive Humby, a British mathematician: “Data is the new oil. Like oil, data is valuable, but if unrefined, it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity. So, must data be broken down, analyzed for it to have value” (Pycoach). Data analytics is a multidisciplinary field where a combination of concepts from mathematics, statistics, and programming are used to analyze and interpret data and the demand for data analysts is growing at a tremendous rate. Companies from industries like e-commerce, finance, insurance, healthcare, education, and sports are deploying data mining techniques to assist managers in the decision-making process (Gupta 222). Most of the service sector companies deploy data mining techniques to boost customer service. It is worth noting that good customer service will lead to better customer retention rates in the long run, which will in turn increase the revenue.

Banks play an important role in the modern economy as they act as a mediator between borrowers and depositors. The main source of revenue for banks is through high interest rates charged on loans. Banks also earn revenue by charging fees from customers for various services. The data generated in the banking sector is highly confidential in nature. Like other companies,

banks can also use customers' financial data to enhance their decision-making process. At present, 12% of banks have started implementing data analytics techniques and 25% of the banks have reported growth (Gupta 222). The main purpose behind implementing various data mining techniques is to increase their credit card sales and to judge the credibility of a customer before lending. Given the volume of data generated in banking and the rigorous and time-consuming data mining process, banks need to ensure that the benefits of implementing data mining techniques outweigh the costs of labor and technology and relative risks.

The term big data is categorized into five types: volume, variety, velocity, value, and veracity. The volume of big data is considered of utmost importance as it is measured in terabytes and the cost of storing the data on cloud platforms is a major challenge faced by data scientists. Raw data is available as structured (or quantitative data) and unstructured (or qualitative data). The variety of data available is crucial at the data extraction and cleansing stage. It is worth noting that data cleansing is the most time-consuming process and subsequent stages such as data visualization and data analysis are dependent on the datasets available after cleansing. Velocity of data refers to the speed at which data is generated, entered, and processed in the system. Value as the name suggests explains the potential of the data. Veracity is best defined as the quality of the data being analyzed (Smaya 3-4).

It is important to understand the pros and cons of data analytics before analyzing its impact on the banking sector. The advantages of implementing data analytics are enhanced decision making, cost reduction, and better customer service. E-commerce companies like Amazon and Netflix are combining data analytics with the concept of personalized marketing to tailor the customers' experience based on historical data. Data from the search results, purchase history, and watch history are cleansed and analyzed to find trends and recommend products and

services based on the user's interest (Le and Liaw 5). Statistics presented in an article imply that Amazon makes 35% of its sales revenue from personalization using data analytics. The ability of e-commerce companies to retain and expand their customer base in the long run with the help of analytics is indeed impressive (Khandelwal). The concepts of demand and supply are applied along with data analytics to find the equilibrium price so that companies can charge the customers with the highest price they are willing to pay for goods and services. Data analytics is used in the healthcare industry to find trends related to the spread of various diseases (Le and Liaw 5).

Although the benefits of data analytics are compelling, there are some drawbacks that companies need to consider before making an investment. Companies need huge server farms and cloud networks to securely store the data online and these investments can affect the profitability of a business in the short run. Most of the data is confidential in nature and thus, companies must ensure the privacy and security of customers' data (Le and Liaw 6). Data leaks affect the goodwill of the company and the customer base which in turn affects the revenue and growth of the company. There is a massive shortage of data analysts in the job market at present. Data mining and data cleansing are extremely time-consuming. Thus, companies must consider the cost of labor and technology before making an investment in analytics.

Data analytics is classified into four types: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. Descriptive analytics is used by data analysts to solve problems based on past happenings. For example: A sudden spike in the sale of electronics between October and December. Diagnostic analytics is used to find answers to various problems raised. For example: Electronics are a popular option for gifting during the holiday season. Predictive analytics is used by data analysts to predict what could happen in the future.

For example: Considering the spike in the sale of electronics during the holiday season in the past, it is highly possible for the same trend to continue in the upcoming holiday season.

Prescriptive analytics is used to project a future course of action. Data analysts can predict the right time to start promotions related to the holiday season and suggest methods to improve inventory control (Cote).

The data analytics process begins at a stage where data scientists and data analysts extract data from a variety of online sources including online databases like company records and websites. The data is securely stored in a cloud-based data warehouse or an online server farm. Data from these sources is cleansed and the information deemed relevant for the study is retained in the database. Unstructured data is converted into structured data and the information is sorted by preference. Errors in the database are also rectified and filtered before moving onto the next process. Data analysts visualize the information based on various parameters using visualization tools like Power BI and Tableau to find trends in the dataset. Based on these trends, data analysts recommend the best options for solving their client's problem and discuss its impact on the performance and growth of the organization (Smaya 5-6).

With regard to the latest trends in the banking sector, digital banking is extremely popular around the globe these days as countries are switching to a cashless economy to prevent financial crimes and curb corruption. Compared to the traditional banking system, customers can avail banking services remotely from the comforts of their homes. Customers can access their bank statements, transfer, and manage funds with ease, apply for debit and credit cards, and avail housing and car loans through net banking (Sandhu 2169). Considering the high volume of transactions made every second, the data generated from these transactions worldwide can be used effectively by data analysts to enhance customers' experience with banks. Data analysts can

assist various levels of management in taking the right decisions on lending funds based on credit history, improving credit card sales, setting ideal interest rates etc. to develop the bank's customer base in the long run.

One of the benefits of implementing data analytics in banking is to ensure security and detect fraud. Considering the confidential nature of data in the banking sector and the increased risk of cyber crimes around the world, researchers are working on finding advanced data mining techniques to protect customers from hackers. The most popular techniques proposed by researchers to prevent credit card fraud and online banking fraud are k-means clustering and the use of neural networks (Hassani 3). Neural networks are a branch in artificial intelligence where computers are programmed to mimic the functioning of the human brain. The computer gains experience in the long run by committing mistakes and rectifying them ("What Is a Neural Network? - Artificial Neural Network Explained). Under k-means clustering, data points are repeatedly assigned to groups. These data points segregate into groups based on similar features (Sharma). Cryptography i.e., the act of securing communication by allowing only the sender and the recipient to read the message (Cryptography Definition) was proposed as one of the solutions in a research study. The data mining techniques of k-mean clustering and neural networks were also used to detect phishing and money laundering cases. Another technique called blog mining was implemented to analyze search results for security risks and protection methods. The data generated was mostly from popular blog sites like the Google search engine. Thus, data mining techniques like k-means clustering, neural networks and blog mining were deemed successful with high rates of accuracy by the researchers for finding insights on fraud detection and ensuring security by offering solutions (Hassani 3).

Banks need assistance with risk management and researchers have tried and tested various algorithms in the past and are working to enhance the process in the future. The main concept involved is to apply data mining techniques to the credit scoring process. Some of the most crucial functions of banks include lending loans, offering overdraft facilities and credit card payment options to their customers. But it is important for the bank to analyze the credibility of the customer using their credit history before lending funds as customers should have the capacity to repay the loan with the interest due on repayment. This step helps banks to reduce the number of customers who file bankruptcy.

Researchers used an algorithm called support vector machine clustering with a German bank's credit scoring dataset and another researcher introduced a sub-type of neural networking called multi layer perceptron (Zarah 155). In a support vector machine algorithm, the data is split into classes through classification and regression models (Pupale). A multilayer perceptron is used to classify data into various input and output layers with a hidden layer in between (Bento). These data mining techniques along with k-mean clustering help banks to flag customers with poor credit scores before lending funds. The credit information sets were changed in a recent study to test another form of data mining called Bayesian hyperparameter optimization. Thus, researchers are working to find additional methods which help banks to perform risk management effectively and to improve the lending mechanism (Zarah 155).

As competition between banks in the industry intensifies, it is vital for banks to build a bond with their customers. To increase revenue and growth, banks need to maintain and grow their customer base in the long run. Customer relationship management is a field which improves the organization's customer retention rate and customer satisfaction levels by analyzing the needs and demands of the market using technology (Chuang 420). Customer relationship

management has been regulated by data mining techniques. Banks play a crucial role in extracting data based on the customer's spending habits. Customer segmentation is a process of splitting customers based on factors like demographics, behavior, and attitude. Researchers accumulated data for analysis from banks on topics like advertising and savings account information to test a combined algorithm and the profits increased tremendously. Data was collected through questionnaires in China for finding trends and researchers found a connection between mobile banking and Chinese demographics. Thus, data analysts can help banks initiate attractive schemes to retain and grow their customer base. Data analysts support banks in improving client satisfaction. Data mining techniques are used effectively to develop trust in clients in Spain (Zarah 156).

Banks have used data mining to flag suspicious activities like a decrease in the number of transactions. Customers are encouraged to invest funds in the account with attractive interest rates or even encouraged to spend funds by offering incentives. These steps help the banks retain their customer base in the long run. Support vector machine clustering and neural network classification were tested on various datasets from banks all over the world to find trends on the number of customers switching to competitors for services and investments. Thus, banks have tested various data mining techniques along with customer relationship management to find trends and improve the quality of service provided to its customers. These techniques were deemed successful in attracting customers from competitors and increasing the activity of existing customers regarding credit card sales, increased spending, improved lending based on trust etc. (Hassani 6-7).

The existing research on data mining techniques like k-mean clustering, neural networks, support vector machine clustering, and neural network classification has been successful in

improving the state of banking. However, compared to other fields in technology like artificial intelligence and machine learning, research on data analytics and data mining techniques is limited at present. Data in banking is extremely confidential and considering the volume of data in the banking sector, it is a rigorous process to hire labor for data processing, mining, and cleansing. The government should collaborate with private organizations to invest in additional research on data mining techniques in e-commerce, healthcare, finance, sports, and other industries.

There is a massive shortage of skilled data analysts in the job market worldwide as the demand for data analysts is growing tremendously and it is exceeding the supply. Considering this massive demand for professionals, the government and other stakeholders must make changes to the education system to introduce an additional pathway in high school with an emphasis on technical and soft skills required for data analytics, machine learning, artificial intelligence, and cloud computing. Universities around the world should introduce majors in these areas with rigorous training and practical exposure so that students enter the workforce with sufficient experience upon graduation.

The return on investment in technology and labor is observed only in the long run. The benefits of implementing data analytics in banking are impressive. However, there are more borrowers than depositors in the economy at present and banks should work on developing customer relationship management to increase their customer retention rate and expand their customer base. Banks should invest in data mining techniques only when additional research work on data mining techniques is made available. Researchers must make convincing arguments in their papers that the benefits of implementing data mining techniques outweigh the huge cost of investment in the long run.

Works Cited

- Bento, Carolina. "Multilayer Perceptron Explained With a Real-Life Example and Python Code: Sentiment Analysis." *Medium*, 5 Jan. 2022, towardsdatascience.com/multilayer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141.
- Cote, Catherine. "4 Types of Data Analytics to Improve Decision-Making." *Business Insights Blog*, 19 Oct. 2021, online.hbs.edu/blog/post/types-of-data-analysis.
- Chuang, Chao Chao, and Fu-Ling Hu. "APPLICATION OF CRM IN BANKING." *Aktual'ni Problemy Ekonomiky*, no. 154, 2014, p. 419–.
- "Cryptography Definition." *www.kaspersky.com*, 9 Feb. 2022, www.kaspersky.com/resource-center/definitions/what-is-cryptography.
- Hassani, "Digitalisation and Big Data Mining in Banking." *Big Data and Cognitive Computing*, vol. 2, no. 3, MDPI, July 2018, p. 18. <https://doi.org/10.3390/bdcc2030018>.
- Khandelwal, Astha. "How Does Amazon and Netflix Personalization Work?" *Blog*, 22 Feb. 2023, vwo.com/blog/deliver-personalized-recommendations-the-amazon-netflix-way.
- Le, Thi Phuong Quynh, and Shu Liang Liaw. "Effects of Pros and Cons of Applying Big Data Analytics to Consumers' Responses in an E-Commerce Context." *Sustainability*, vol. 9, no. 5, MDPI, May 2017, p. 798. <https://doi.org/10.3390/su9050798>.
- Pupale, Rushikesh. "Support Vector Machines (SVM) — an Overview - Towards Data Science." *Medium*, 9 Mar. 2022, towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989.

PyCoach, “Is Data the New Oil of the 21st Century or Just an Overrated Asset?” *Medium*, 21 July 2022, towardsdatascience.com/is-data-the-new-oil-of-the-21st-century-or-just-an-overrated-asset-1dbb05b8ccdf.

Sandhu, Supreet, and Sangeeta Arora. “Customers’ Usage Behaviour of E- banking Services: Interplay of Electronic Banking and Traditional Banking.” *International Journal of Finance & Economics*, vol. 27, no. 2, Wiley-Blackwell, Oct. 2020, pp. 2169–81. <https://doi.org/10.1002/ijfe.2266>.

Sharma, Natasha. “K-Means Clustering Explained.” *neptune.ai*, 20 Feb. 2023, neptune.ai/blog/k-means-clustering.

Smaya, “The Influence of Big Data Analytics in the Industry.” *OALib*, vol. 09, no. 02, Scientific Research Publishing, Dec. 2021, pp. 1–12. <https://doi.org/10.4236/oalib.1108383>.

“What Is a Neural Network? - Artificial Neural Network Explained - AWS.” *Amazon Web Services, Inc.*, aws.amazon.com/what-is/neural-network/#:~:text=A%20neural%20network%20is%20a,that%20resembles%20the%20human%20brain.

Zarah, Ameliya Eleanor, et al. “Big Data Analytics and Its Impact on Efficient Banking in Europe-Asia.” *Business & IT*, vol. XII, no. 1, Czech Technical University in Prague, Jan. 2022, pp. 152–59. <https://doi.org/10.14311/bit.2022.01.18>.