UTILIZATION OF DATA MINING AND MACHINE LEARNING IN DIGITAL AND ELECTRONIC PAYMENTS IN BANKS

Alban Rakipi *, Olimpjon Shurdi **, Julian Imami **

* Corresponding author, Faculty of Information Technology, Polytechnic University of Tirana, Tirana, Albania
** Faculty of Information Technology, Polytechnic University of Tirana, Tirana, Albania

Abstract

The purpose of the paper is to study and measure the usability level of new technologies such as machine learning (ML) and data mining (DM) in the banking sector. An overview of the current situation in the international markets and a deeper look at the Albanian one is given. The research methodology includes qualitative and quantitative methods. This mixed approach offers the possibility of obtaining detailed answers and allows statistical analysis to draw patterns or trends related to the subject of the study. This research topic is important since the studies in this field in the regional context are few and this topic has not been extensively studied before. New technologies and their advantages are of great importance to the banking sector. This study identifies prevalent ML approaches currently and is expected to be used in the near future. An online survey is done on a network of professionals in the banking and financial sector. The findings show that the impact of ML and DM in these organizations, currently is at an average level. The results of this study can be a useful indicator for the stakeholders of the banking sector to implement new digital technologies such as ML and DM.

Keywords: Digital Banking, Mobile Banking, Data Mining, Machine Learning, Big Data

1. INTRODUCTION

The innovation in digital banking is supported by adopting potential digital technology, such as artificial intelligence (AI), big data, blockchain, cloud, and the Internet of Things for process automation and more intelligent banking services (Indriasari et al., 2022).

The paper's objective is to bring light to the utilization, adoption, and knowledge of people on technologies, practices, and methods of data mining (DM) and machine learning (ML) in the banking sector (Chitra & Subashini, 2013; Damrong sak methee & Neagoe, 2017; Hassani et al., 2018) and financial sector in Albania, but, at the same time, comparing and analyzing the situation internationally.

Many industries are changing due to digitalization, as well as how businesses and customers interact. The banking sector has been notably affected by the digital revolution, with customers increasingly using digital banking (online and mobile) as one of their most important channels (Carbo-Valverde et al., 2020). The main advantage of digital banking services is their ease and function in replacing the role of branch offices (Musyaffi et al., 2022). The banks need to identify the challenges before digital banking initiatives to prevent failure in digital banking innovation (Indriasari et al., 2022).
The data for this study was collected from a survey where respondents provided their email addresses. Acquired data from second-level banks in Albania was filtered off for sensitive info and attribute names were hidden.

The study’s major research questions are:

**RQ1:** Are data mining and machine learning techniques being used in the Albanian banking industry?

**RQ2:** What is their influence and how strong is the feeling associated with it, if used?

In addition, this research aims to identify the most common machine learning approaches that are now utilized or are expected to be used in the near future.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the research methodology. Section 4 analyses the present status of the study. Section 5 presents the main results of the research work. Section 6 discusses the findings and, finally, Section 7 concludes the paper.

## 2. LITERATURE REVIEW

Digital banking plays a significant role as an enabler of cashless transactions in the economic crisis caused by the COVID-19 pandemic (Indriasari et al., 2022).

Machine learning and data mining are an extremely big and important part of “big data” and in the past two years have played a major role, especially in the digital transformation of the banks (Doerr et al., 2021).

The increase in the demand and the popularity of using ML/DM in the banking sector is a must for the post-pandemic situation which is considered a great increase and boost in online services.

Recent studies (Pagani & Pardo, 2017; Poustchi & Dehnhurt, 2018) suggest that although tools and computable power have increased over the years, utilization of data mining techniques and machine learning, although increased over time, are still not at the expected level and impact on our daily lives.

Both data mining and machine learning have seen greater adoption over the years in many industries but adoption in certain ones like banking has seen a slow spread (Hassani et al., 2018). Whether automotive, airline, or pharmaceutical industries all have, to some extent, adopted methods to classify their customers and production, to use AI to predict future peak needs, adjust pricing, and many other fields where the main goal is to leverage profits using up to date methods of DM and ML (Damrongkasmethee & Neagoe, 2017).

Hassani et al. (2018) concluded that the banking sector mainly adopts DM techniques for the following purposes: security and fraud detection, risk management and investment banking, and customer relationship management (CRM). DM techniques have been widely applied in banking for marketing and CRM-related purposes, such as customer profiling, customer segmentation, and cross/up-selling. These tools help banks to have a broader understanding of their customers and predict their behaviors accurately.

Al-Smadi et al. (2023) underline that combining the power of data mining and machine learning with blockchain technology offers enhanced security, transparency, and efficiency in payment transactions.

Doerr et al. (2021) underlined the need for ML in central banks because the central bank usually has more data than a single bank but, in the meantime, more responsibility to understand and predict market and client behavior.

The results of Carbo-Valverde et al. (2020) suggest that banks should address the digital transformation of their customers by segmenting them according to their revealed preferences and offering them personalized digital services.

With the emerging technologies in digital currencies but also increasing open-source solutions to many information technology (IT) problems combined with the increase of ease qualification of IT professionals in developed countries has resulted in poorer countries in development having the ability to implement modern strategies in IT and IT infrastructure easier than the developed countries (Pagani & Pardo, 2017).

Musaffi et al. (2022) concluded that the high number of transactions through branch offices with large transactions shows that consumers are still hesitant when making transactions through digital banking services.

Kaur et al. (2021) proposed a model that highlights the impact of in-branch customer engagement on their intention to adopt digital banking channels in India. Migration of customers to digital banking channels needs to develop a branch transformation strategy with a focus on a customer-centric approach. These issues raise the urgent need to focus on improving the in-branch practices of banks to convince customers to migrate to digital banking channels.

To discover critical accounting variables that could explain bank failures, Mulyadi and Anwar (2023) used automated machine learning techniques and the local interpretable model-agnostic explanations (LIME) method to find five critical accounting factors. The findings’ importance is in evaluating financial institutions, providing useful insights for stakeholders, decision-makers, and perspectives for future research.

Ha et al. (2023) employ ML algorithms to measure the probability of financial distress.

Plotnikova et al. (2019) discuss the need for and give a survey of data mining and advanced data analytics approaches, with a focus on their use in the banking industry. One hundred and two (102) papers were collected and analyzed using a systematic literature analysis in order to answer three research questions: For what goals are data mining approaches utilized in the banking domain? How are they used ("as-is" vs. modified)? What are the objectives of adaptations? A prevalent practice in the banking industry of using data mining approaches “as-is” to solve CRM and risk management business challenges was discovered.

One of the most used cases of AI in many businesses is process management and process automation. This can streamline behind-the-scenes processes and significantly improve processes. Repetitive tasks, easy tasks that are through daily work data-driven can be shortened and automated.

The financial services sector integrates AI to analyze massive data, track scams by detecting suspicious transactions, interact electronically with clients, and explore a number of other key functions. Many of the key aspects include electronic trading,
portfolio structure and optimization, robotic consulting services, automated customer services, business forecasting, and other forms of data-intensive analysis (Donepudi, 2019).

It is essential for banking institutions to prioritize AI that will bring significant changes to the specific areas of the data strategy. The process required by the data management includes providing an original set of historical data to teach the ML processing strategies, and continuous data insertion to improve and also train the accuracy of the models (Parne, 2021).

Many theories attempt to explain the development of these novel technologies, as well as the relationship between the corporation and the customer. For describing how individuals accept and embrace new technology in the context of banking, the technology acceptance model (TAM) (Davis et al., 1989) and its later versions (TAM2, TAM3) have gained popularity. According to the TAM model, which is founded on the theories of reasonable action (TRA) (Ajzen & Fishbein, 1975) and planned behavior (TPB) (Ajzen, 1985, 1991), the acceptance of new technologies depends on how useful and simple they are for customers to use.

Big data is not accepted at a satisfactory level because banks have had mainly old legacy systems which are built to serve clients and its core services for decades. The reasons include complex updates and developments that require heavy investments. Because of this issue, the risk of low adoption of ML and DM falls dramatically due to lack of interconnection, poor or missing data, and wrong transformation of data. Not only system issues but also, according to interviews, it seems that customer data itself was not a topic of great importance in the banking system rather than their financial data (Carbó-Valverde et al., 2020).

Banks are risk-oriented, meaning changes and developments take a long time due to extensive testing and fear of trying new technologies. This is why technology growth is slow due to risk orientation.

Digital banking research trend experienced significant growth from 2015 to 2019, but slowed down from 2020 to 2021 due to fintech and startups posing more appealing challenges (Indriasari et al., 2022). The study of digital banking is an important investigation topic in the era of innovation, and many researchers from different theoretical views have investigated it. Rarely a study in Albania has given attention to the utilization and adoption level of new technologies like ML and DM in the banking sector. The present study tries to close this research gap by answering the formulated research questions.

3. RESEARCH METHODOLOGY

The research has been conducted through an online questionnaire. Both qualitative and quantitative questions were included in it.

Mixed methods research like the one used in this study includes collecting and analyzing both qualitative and quantitative data. The goal of questionnaire is to elicit detailed insights and experiences from the persons who answered the questionnaire. Quantitative inquiries, on the other hand, are designed to collect numerical data that may be statistically examined to uncover patterns, connections, or trends. By capturing both subjective and objective viewpoints, this approach can capitalize on the strengths of a more thorough and holistic understanding of the research issue.

Qualitative questions include open-ended responses related to the industry and the sector where the respondents work. It also covers the subjective aspect of the employees' experiences.

Quantitative questions have been used to collect numerical data and then to express them in a graphical view and analyze them.

Closed-ended questions with five levels of choice, ranging from low to high, are employed in this scenario as well as multiple-choice questions.

An alternative method to collect data might have been through group or individual interviews, focus groups, case studies, or content analysis.

To get more information about AI and DM, a questionnaire with ten questions was designed. The first part of the questionnaire contains some informative data about the subjects who participated in the study. The second part of the survey inquires about the frequency of data analysis, data source, use of ML or DM in daily basis work, impact of ML or DM in company, strength of impact if organization starts using them, current impact of ML or DM, and preference for unsupervised learning, supervised learning, or reinforced learning in use or planned use in the organization.

A questionnaire was conducted to collect the responses from the respondents between February 2022 and April 2022. The goal is to get deeper insights into how bank employees perceive and think about the usability of new technologies (ML and DM). The questionnaire is shared with more than sixty people who work in public and private sector banks, but also a few in other fields. It is worth mentioning that most of the participants are from banking and financial institutions and from central banks (above 65%), which makes this questionnaire truly relevant to the study. In summary: the data collection is done through a questionnaire, with a total sample of 63 participants, with a scope of Albania. For the analysis purpose, exploratory data analysis (EDA) and R programming are used.

4. DIGITAL BANKING IN ALBANIA

In this section, the present situation in Albanian banking and tech companies that utilize modern approaches in big data and have incorporated in their company ML/DM methods is analyzed.

COVID-19 has revolutionized the banking industry and digital banking has become increasingly important due to the lockdown and fear of infection. People are increasingly turning to digital services.

Recently, there has been a growing need for international policy cooperation, especially among public and private authorities in the financial sector to facilitate the use of payment data and promote innovative technological solutions (Cornelli et al., 2022).

The true essence of digital banking is to allow customers to have the freedom of doing banking transactions from anywhere even outside Albania without coming to a branch. This translates into the more digital behavior of the client.
The challenge for banks is to understand the customer's expectations and the reasons they choose (or not) to connect virtually with their bank (Filotto et al., 2021). In order to adapt how they connect with consumers, manage middle and back office operations, stay competitive, and be ready for the future, banks need to modify their business strategy (Kitsios et al., 2021).

Digital banking in Albania is being explored through MI, and its growth over time is being compared. Tables 1 and 2 illustrate the exponential growth of both Internet and mobile banking.

Mobile banking has seen growth a couple of years later than digital banking was introduced but the adoption rate is much higher for mobile banking. It is impressive to see that in just 3 years, transfers in the digital banking industry have spiked from 2.5 million to 5 million in the year 2021. It is very impressive to see that the ratio between digital transfers to physical transfers has come from 4% to 46% over the past 10 years with a high peak around the COVID-19 pandemic (Bank of Albania, 2022).

As a result, banks in Albania are leveraging a variety of digital channels to the fullest extent possible, with a particular emphasis on mobile banking, e-banking, and self-service machines like ATMs. As illustrated in Tables 1 and 2, the increase in digital transfers is a clear indicator of banks investing heavily in self-service and digital banking channels in recent years.

The advantage of the mobile banking application is that the client may access their accounts, get information in all accounts, make transactions, and make domestic transfers 24 hours a day, 7 days a week. Also, for some special type of transfer, the client will be notified with an SMS from the bank. The most important is that during this period banks have developed projects aiming to enhance the security of banking transactions and innovation in a banking environment.

### Table 1. Digital banking transfers from 2012 to 2021 in Albania

<table>
<thead>
<tr>
<th>Year</th>
<th>Internet Banking</th>
<th>Mobile Banking</th>
<th>Total Digital (Internet &amp; Mobile)</th>
<th>Year-over-year growth (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>263,187</td>
<td></td>
<td>263,187</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>327,648</td>
<td></td>
<td>327,648</td>
<td>36.6</td>
</tr>
<tr>
<td>2014</td>
<td>802,282</td>
<td>25,081</td>
<td>827,363</td>
<td>156.4</td>
</tr>
<tr>
<td>2015</td>
<td>934,309</td>
<td>27,031</td>
<td>961,340</td>
<td>16.3</td>
</tr>
<tr>
<td>2016</td>
<td>1,236,879</td>
<td>25,260</td>
<td>1,262,139</td>
<td>33.2</td>
</tr>
<tr>
<td>2017</td>
<td>1,520,091</td>
<td>178,720</td>
<td>1,698,811</td>
<td>32.5</td>
</tr>
<tr>
<td>2018</td>
<td>2,034,466</td>
<td>470,475</td>
<td>2,504,941</td>
<td>47.5</td>
</tr>
<tr>
<td>2019</td>
<td>2,623,341</td>
<td>734,440</td>
<td>3,357,781</td>
<td>34.0</td>
</tr>
<tr>
<td>2020</td>
<td>2,781,303</td>
<td>1,217,647</td>
<td>3,999,950</td>
<td>19.1</td>
</tr>
<tr>
<td>2021</td>
<td>3,268,138</td>
<td>1,787,249</td>
<td>5,055,387</td>
<td>26.4</td>
</tr>
</tbody>
</table>


According to the World Bank Group (Ungerer et al., 2020), the following could be done to increase e-commerce transactions in Albania:

- Basic payment accounts — a government initiative to assist financially vulnerable groups, for example, low-income persons, rural residents, pensioners, out-of-labor forces, etc.
- Ensure information and education on how to activate cards for online purchases.
- Roll out a basic payment card through the Albania Post, including for online shopping.
- Implement a financial education campaign specifically on the use of electronic payments for online shopping.
- Rigorously follow up on cases of online payment fraud and identity theft.

Another important aspect to mention is fiscalization, which means that all issued cash and non-cash invoices and accompanying invoices shall be recorded in real-time to the tax authorities. The system has entered into force in Albania since summer 2021 and is currently running to more than half of the businesses in Albania.

Fiscalization makes sure that invoices are reported in real time and validated in the tax office. This includes e-invoicing for non-cash transactions (business-to-business — B2B; business-to-consumer — B2C).

The new requirement will increase the data generated for payments in Albania, which were previously manual and paper-based. Integration with banks is also mandatory in this process where the bank is currently processing B2B and B2C invoices and reporting online to the tax authorities.

Fiscalization is a complex project in Albania, but it has been adopted by many players. It brings digitalization and benefits to reduce the informal economy and trace payments between businesses. Digital developments in banking around the world have also proven to impact the services offered in Albania.

After the COVID-19 pandemic, the acceleration in offering more and more services through web or mobile banking has increased significantly. The pandemic showed that digitalization is not a luxury but a necessity.

### Table 2. Digital banking versus paper-based transfers in Albania

<table>
<thead>
<tr>
<th>Transfer initiated from</th>
<th>2012</th>
<th>2021</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper (in branch)</td>
<td>6,376,269</td>
<td>5,901,549</td>
</tr>
<tr>
<td>Electronically</td>
<td>238,213</td>
<td>312,9064</td>
</tr>
<tr>
<td>Total</td>
<td>6,614,484</td>
<td>11,014,013</td>
</tr>
</tbody>
</table>

5. RESULTS

The obtained results of the questionnaire are depicted in Figure 1 to Figure 9. Regarding the first question, about the institution where the respondents work, most of them stated that they work in a bank, as shown in Figure 1. Although some are reported as audits, for example, private companies, or government they are linked to the banking or financial system, an estimate of 65%.

**Figure 1.** The distribution of the respondents according to the work institution

![Pie chart showing the distribution of respondents based on work institution.](image)

**Note:** NGO is for “non-governmental organization”.

When the subjects were asked in which field or department they work, 17% stated that they work in IT software development, and 17% worked in finance, accounting, and reporting, as shown in Figure 2. In total, seeing the respondent’s profiles, around half of them are working in the IT sector and mostly in the banking profile.

**Figure 2.** The distribution of the respondents according to their departments

![Pie chart showing the distribution of respondents based on departments.](image)

**Figure 3.** The familiarity of the respondents with the technical terms

<table>
<thead>
<tr>
<th>Technical Terms</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>None of them</td>
<td>17%</td>
</tr>
<tr>
<td>Neural networks</td>
<td>22%</td>
</tr>
<tr>
<td>Artificial intelligence</td>
<td>69%</td>
</tr>
<tr>
<td>Deep learning</td>
<td>28%</td>
</tr>
<tr>
<td>Data mining</td>
<td>52%</td>
</tr>
<tr>
<td>Machine learning</td>
<td>59%</td>
</tr>
<tr>
<td>Big data</td>
<td>66%</td>
</tr>
</tbody>
</table>
When the subjects were asked about the most familiar term, the study found that 69% of participants identified “artificial intelligence” as the most familiar term, as shown in Figure 3.

Figure 4. The frequency of using data for analysis

Regarding the data source, in Figure 5, the subjects that participated in the study stated that the data source is the main system of the company, about 48% of the subjects. As was expected 14% and around 33% are using either a data warehouse or locally generated data.

When the subjects are asked if they use ML or DM in the work, about 76% say “No”, as shown in Figure 6. This is, in fact, a result showing low acceptance in the industry.

Figure 5. The share of data sources

Figure 6. The usability of machine learning and data mining

In Figure 4, most of the respondents, about 38%, stated that they use data for analysis every day, multiple times. Only less than a quarter never use data or use it not so frequently.

Figure 7 depicts the results where the subjects are asked to scale from 1 (low) to 5 (high) how big is or would be the impact of ML or DM in your organization — about 33% rate it with “4”. Also, regarding the impact of ML or DM and how strong it will be if the organization starts using it now, about 33% of the respondents rated with “4”.

Figure 7.
Figure 7. The impact of machine learning and data mining on the organization

Figure 8. How strong is the impact of machine learning and data mining in the organization

For the respondents that gave positive feedback to question 4, they use ML and DM in their company; a rate of 3 for the strong impact of ML and DM in the organization was given.

The last question was whether they use/plan to use ML or DM in the organization; then which one they use/plan to use more; and about 64% stated supervised learning (classification, regression), as shown in Figure 9.

Table 3 summarizes the perceived use, in terms of average and standard deviation for three of the main questions of the questionnaire, where it is noticeable the future impact is important.

Table 3. Statistical results for the provided answers

<table>
<thead>
<tr>
<th>Question</th>
<th>Average (Mean)</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>How strong is the impact of ML/DM in your organization, currently?</td>
<td>2.57</td>
<td>1.28</td>
</tr>
<tr>
<td>How strong do you believe the impact of ML/DM will be if the organization did not previously use them and now does?</td>
<td>3.38</td>
<td>1.24</td>
</tr>
<tr>
<td>How big is or would be the impact of ML/DM in your organization?</td>
<td>3.40</td>
<td>1.23</td>
</tr>
</tbody>
</table>
6. DISCUSSION

Big data adoption in Albania has started in the last couple of years due to emerging technologies, increased data, digital services, and increased systems. This is due to the increase in data, digital services, and systems.

Not only customer operations have increased but also employee operation and system-generated logs now are doubled or tripled compared to past years and need to be addressed into data polices and strategies within the bank.

In all aspects of digital banking, big data, machine learning, and data mining are processes that although new in Albania will be in demand and start being used by the banks. This is a path that needs to be followed in order to cope with the increased data generated from many services, aspects, and solutions the banking industry is working with.

This study collected data from a survey and acquired data from second-level banks and central banks in Albania. The data was filtered off for sensitive info and attribute names were hidden. The challenge was to provide insightful and real-world problem-solving, while also protecting sensitive info and protecting the data.

The main contributions of this paper are:
- conducting an analysis of the digitalization situation of the banking sector in Albania and comparing it in the regional and global context;
- offering a contribution to the existing literature through the results obtained from the study;
- providing a more complete picture of the adoption of innovative technologies such as machine learning and data mining as an important part of digitalization processes;
- these findings could be helpful for policymakers in developing effective strategies to advance financial digitalization.

Regarding the impact that ML and DM have on the financial sector, the findings show that it is at an average level. The results show that if these institutions start using ML/DM or plan to use it in the future, the impact is rated on average with around 3.4 points out of 5, which shows the importance of adopting these technologies in the banking sector in Albania.

It is important to analyze the diversity of banking channels, in particular, the key factors in the implementation of a multi-channel distribution system, customer relationship management through a system to provide a universe of data/information to customers, to collect and provide ongoing analysis of needs, requirements, and preferences, elements which are constantly evolving.

7. CONCLUSION

Banks in Albania now see the need to make policy changes and budgeting to start adopting and investing in new technologies that help perform data mining and machine learning to get insight into customers and serve better them.

The growing use of data mining and machine learning techniques in the banking industry has far-reaching advantages. In the financial sector, ML algorithms improve decision-making processes by predicting unanticipated events, offering the right banking products and services to customers, etc. The study's implications relate to the penetration of these technologies in the Albanian banking industry.

The questionnaire results indicate that big data offers significant potential for growth, with numerous reasons and ways to expand in this direction.

The first conclusion based on the questionnaire of our research is that in banks and financial institutions, data analysis is done frequently from several times a week to several times a day.

Second, only a quarter of respondents use DM or ML for daily tasks according to questionnaires but even less when discussed face to face.

Third, the currently perceived impact and the future of those models is on average and people still believe that big data needs to be advanced in some fields and still is not accurate in all matters.

Fourth, most data mining is basic, and on the level of EDA, which is mostly employed in everyday life.

Lastly, we can state that new technologies and new regulations will force banks to make structural and strategic changes to the core business they offer since digital payments are coming a long way and new solutions are entering aggressively in the market.

Future implications on how and why the ML methodologies will see mass adoption in Albania are increased card and ATM transactions, PSD2 (Payment Service Directive 2), compliance and open banking, increased mobile banking transactions, and new emerging technologies in banking and payments.

ML and DM processes are essential in coping with the already growing data needs that the industry is already facing in the last five years. Although banks are running into legacy systems and having lots of obstacles it is obvious that the growth of data is becoming huge.

The limitations of this research work are related to the sensitivity of the data provided by the subjects of the questionnaire. In order to obtain more substantial conclusions, we think that scalability is needed in receiving responses from a higher number of users of the sector and the results should be correlated with the stakeholders’ approach to implement and widely embrace new technologies in the banking system.

Future work is related to the study of the most advanced methods and their implementation in order to have a more comprehensive panorama of the use and adaptation of innovative technologies and how can ML/DM utilized for improving customer experience.

REFERENCES


