AI-Driven Customer Behavior Analysis in Banking

By **Dr. Priya Rajagopal**

Professor of Mechanical Engineering, Indian Institute of Technology Bombay (IIT Bombay)

1. Introduction

In the modern world where feedback is the most valuable part, data generation for customer feedback is an everyday job. Users should not underestimate the importance of having feedback from bank customers. The banking sector is one of the fields where Customer Relationship Management (CRM) is necessary. CRM is the foundation for customer satisfaction, retention, and loyalty, and a prerequisite for successful relationship marketing. The success of a business in the banking sector is mostly connected to the satisfaction of its customers, as customers are the most important resource, and they should be aware of their value to the business. This study aims to provide banking sector firms with helpful insights to create better business plans while using new data analysis techniques by creating recently developed risk indexes. The age of Big Data starts when part of the overall data is structured while the majority remains unstructured. The largest portion of this data is generated from users, and banking will not only collect feedback based on the products they offer. As models show trends in the big data sector, it is evident that these models can be used to increase data quality and business outputs. The most expanded areas where models are being applied are banking, fintech, and customer feedback sentiment analysis.

1.1. Background and Significance of AI in Banking

In the financial industry 4.0 era, the implementation of artificial intelligence technology is not only a natural evolution but also a compulsion. This is a result of the increasing operational challenges brought about by the urgent demand for real-time decision support, risk management, fraud detection, workflow automation, customer analytics, and improved efficiency, among others. Financial institutions that neglect this reality will increasingly be subject to higher operational costs, which will manifest as the bank becoming less agile, larger labor operations, and increased risk. Furthermore, to provide customer-centric banking automation that meets the customer's needs for convenience, financial institutions will require customer-centric technologies. Artificial intelligence technologies are expected to play a significant role in this financial digitization as key technologies in these digitization transformations, alongside the cloud and API revolution. Applications will be driven by every sector, ranging from front-end intuition to back-end operations. There are four main categories where AI is implemented in banking sectors as follows.

Over the decades, customer behavior has not been well understood by the financial sector. This is not only particularly risky but also forgoing significant opportunities to expand and deepen the relationship with the customer. This lack of understanding is evident to customers themselves. Financial institutions in retail and business sectors can hardly understand customers across the product silos. These silos are exacerbated by the bank's data being scattered due to mergers and acquisitions. Furthermore, digitization has driven customers to move from branch interaction to the internet of things and to the internet of everything, thus making it difficult to understand the reasons behind the patterns of customer behavior. Financial customers want personalized service based on their digital footprints across the product silos of the bank for a variety of problems, not just purchase recommendations. Banks need to begin to understand the strength of the emotional relationship with their customers.

2. Foundations of Customer Behavior Analysis

This section introduces the specifics of customer behavior analysis within the banking industry using transactional banking data. We start exploring the features and limitations in terms of the utilized data. After that, we introduce the most widely used features in predictive modeling that have been especially developed for customer-level analyses. In the last part, we discuss the potential business value that lies behind the identification of meaningful customer behavior patterns.

Transactional Data in Banking The most interesting data used in the banking industry remains the transaction data, which are significant in the area of customer-centered analyses, branch network balancing, and credit risk evaluation. In the current work, we concentrate on customer-level analyses. Transaction details contain a rich body of information about a customer and are considered to provide better insight and a more detailed understanding of customer behavior, comparing first and second generation features discussed below. The time of transactions, merchants' category codes identifying merchants through the business type, as well as the amounts, are the main sources.

Customer Level Data Characteristics While discussing customer behavior modeling, the data used are mainly transformed from transactional to customer level. Consequently, generated analytical customer-level features capture information related to the transactional data only. Since a single customer can generate a huge amount of data, it is useful to transform it to an aggregated form, introducing statistics such as count, mean, median, standard deviation, minimum, and maximum. It is also recommended to separate basic raw data and cumulative data, such as the number of purchased coffees, the total amount spent on coffee, and the average transaction cost. Other features consider temporal aspects such as month-to-month differences or coefficients of time trends like the average number of transactions performed per month. Moreover, the number of transactions in the evening versus the number of transactions carried out in the morning uncovers whether a client buys a morning coffee. Finally, customers can be characterized in terms of frequency, recency, and monetary value.

2.1. Key Concepts and Theories in Customer Behavior Analysis

Two fundamental concepts are needed to understand customer behavior analysis: customer relationship management and data analytics. This paper addresses some key concepts of customer behavior analysis in the banking context, which is governed by data protection regulations. The data protection framework in the UK also defines the obligations of organizations in the banking space. Several concepts relevant to data protection are linked to the customer behavior analysis process, including explicit consent, legitimate grounds for processing, data minimization, data accuracy, data retention, data deletion, data privacy, software systems, analytics resources provisioning, audit trails, and data protection impact assessments. These concepts are addressed in what follows before the specific topic of customer behavior analysis in the banking sector is rigorously addressed.

Explicit consent: Customer permission to process personal data is an important principle. After a customer has been informed of the proposed use of data, their consent has to be obtained for any personal data processing. AI and robotics have a significant role to play in how organizations seek consent from individuals. These tools enable personalized consent dialogues to be generated. The actual consent of the user can be processed using more effective voice recognition tools. Automation in consent management can also be helpful. In many situations, the capture of such explicit customer permission is not straightforward. Assistance is needed to enable the customer to understand what they are being asked about and to influence the form and the shape of their consent. It is a public policy to require that such explicit consent in this sort of situation be more empowered, indeed as part of data protection regulations, so that the customer could remain firmly in control of what they want. The regulations do not specify specific methods or technologies for providing extra authentic consent, particularly with regard to customer behavior and choices around these areas are reflected. Throughout our study, figurative consent constantly refers to these additional factors.

3. Machine Learning Techniques in Customer Behavior Analysis

The customer behavior in banking analysis is a complex challenge that, by nature, entails a large number of features, a lack of labeled data, and ambiguity in identifying and solving problems, requiring models robust enough to capture these kinds of situations with their respective variations. The use of several machine learning techniques aims to maximize the results obtained. These techniques include Random Forest, Gradient Boosting Machine, Gradient Boosting with Elastic Net regularization, XGBoost, and neural networks, as a portfolio with many customers and relationships makes traditional modeling methods less useful. Regarding the XGBoost or Gradient Boosting with Elastic Net regularization techniques, credit-risk scorecards derived from these methods demonstrated a higher capability of prediction than traditional logistic regression, characterizing the most powerful models with the capacity to interpret hundreds of attributes associated with a particular behavior. Additionally, XGBoost allowed more freedom to prioritize high-cost customers and decrease customer churn, presenting higher flexibility to optimize the trade-off between sensitivity and specificity. On the other hand, customer risk analysis can show how customers are going to behave in the coming months. This analysis is very important since a credit line increase can decrease customer churn, and a strategy of recirculation based on specific communication can change a customer's behavior to be satisfied with the bank. The Cox model, Logit, Ridge, and LASSO regularized contrasts, RF, GBM, XGBoost, and SVM models can be applied to this problem. Additionally, neural networks are generally used to solve attrition estimation using credit card data. This technique is particularly effective in capturing non-literal patterns due to the reality of operational costs, which means an optimizing process has more value.

3.1. Supervised Learning Algorithms

In general, supervised learning algorithms are used in modeling customer preferences. The goal of a supervised learning algorithm is to find a mapping from input vectors to a given output vector. The input vectors are observed values from some random vector (features), and the output belongs to a set of known class labels or is otherwise well defined. The training of the model is based on some examples, where the desired output is known. Supervised learning typically requires a careful selection of variables, data cleaning, and data coding techniques. In addition, supervised learning algorithms result in a relatively high level of model interpretability, which is crucial from a regulatory perspective. The method of supervised learning is used in robotics, automated NLP tools, or in CRM for customer segmentation, for example, in clustering customers based on their behavior or risk assessment.

Decision trees, random forests, and gradient boosting machines are the standard algorithms for classification and regression-based supervised learning. Also, the Bayesian method has been quite popular for solving many supervised tasks, especially text categorization tasks, as it has been shown to provide reasonably good solutions with minimal or possibly no tuning at all. In contrast, the performance of algorithms for predictive modeling, particularly complex ones, depends on the members and the interactions of the membership of the categories.

4. Applications of AI in Banking

AI is still in the embryonic stage in financial services, particularly in banking, but is expected to develop fast. Although early AI adopters are particularly concentrated in the technology sector, a significant proportion of AI investments are also being made into financial services. AI offers huge potential in many applications throughout the banking value chain, from customer-facing applications such as chatbots and personal money management to loan origination and credit scoring measures, to efficient processing and streamlining middle and back-office processes.

In loan origination and in other types of credit scoring, models such as regressions, decision trees, ensemble techniques, neural networks, and XGBoost can be utilized. Loan origination tries to advise on the credit risk of a potential loan but often needs supplementary documents, which vary across applications for proof of legitimacy, such as identity, income, and address. Such documents can include passports, identity cards, utility bills, pay stubs, tax documents, business contracts, and also statements from bank information. The process could take a few days, as it needs to handle large, non-scalable manual work while also requiring third-party tools to monitor and report criminals and terrorists. AI can bring rapid improvements in this area.

4.1. Personalized Marketing and Product Recommendations

One of the main purposes of customer behavior analysis in marketing is product recommendation. Effective product recommendation can help banks retain customers, expand user relationships, and improve transaction frequency and trading volume. Meanwhile, personalized marketing is also another essential module that accompanies customer behavior analysis. Analyzing customer behavior data helps marketers understand customers' needs, discover potential users, understand customer preferences, and simply improve marketing effectiveness. AI-driven customer behavior analysis can calculate the personal preferences of a bank's customer and make relevant customer behavior recommendations, enabling banks to discover more business opportunities in the market.

The rule-based recommendation that product marketing staff on some social platforms mainly relies on their own experience to push customers or the products they want has the characteristics of relying on stereotype thinking. AI can help financial institutions and banks build personalized marketing for customers. Machine learning becomes increasingly mature and can accurately analyze and understand customer needs and preferences. Using machine learning can enable banks to establish their own personalized product library, which is tailored for each customer. The recommended products produce personalized pushes for the customers, which helps the bank create precise product recommendations.

5. Challenges and Ethical Considerations

Observing and analyzing individuals' patterns of interactions using sophisticated computer algorithms raises important privacy issues. Yet state-of-the-art experimental economic research implies that individuals' sole wish for privacy appears to correspond to recognition of their intentions, and for most interactions, individuals do not have to worry about any harm done to transparency. Apart from such implications, however, this argument for disclosure and low-cost data availability cannot serve as an automatic guideline for the development and implementation of privacy models using AI in predictive algorithms. Bank clients do not only have privacy interests, but safety and security concerns as well.

In recent months, numerous prominent companies have had their clients' confidential data hacked and exploited, and not only banks and fintech companies, with their access to and holding of valuable and confidential information about their private and company customers, have an increased responsibility to attend to the integrity of and provide the security necessities for this private information.

The level of customer service that modern clients anticipate from their banks implies that banks hold large amounts of data about their clients with whom they have confidential relationships. Regardless of the shifting dynamics of banking services and the industry-driven self-regulatory protection standards development, the client-bank relationship is and will remain sensitive and based on trust and confidentiality.

5.1. Data Privacy and Security Concerns

With the development of AI technologies, many customers are starting to refrain from sharing their personal data with businesses because they are concerned about privacy and security. The majority believe that businesses mishandle their personal data. Businesses need to assure all customers that these concerns are properly addressed regarding privacy and data security. Many studies on personal data privacy and security have already been conducted across different research disciplines. The challenge associated with the use of AI technologies in banking can create more privacy and security concerns than before. Therefore, it is believed that upcoming practices and regulatory requirements should help address these concerns and also better align with customer expectations. This can be achieved through persistent dialogue, participation, and engagement with regulators.

Regarding data, legal or compliance measures related to the use of customer data are required to address business challenges and establish an open data framework for increasing AI transparency and accountability. Control of the growing complexity due to AI is necessary for businesses to retain control of their technology. Legal and compliance teams should assess risk, while governance should monitor algorithms and results. Businesses need to be able to understand, manage, and take responsibility for their AI usage. They also need to manage the risk of providing contextual explanations. AI actions that are registered and logged should also be referred to the governance function.

6. Future DIrection

AI technologies, including machine learning and deep learning, are poised to act as tools for the development of systems that aid banking customer analysis for improved business decision-making. The widespread deployment of data mining methods for customer analysis and business intelligence systems in the industry directs the development of parsimonious models because simpler models are more transparent to bank managerial policymakers. Future research can be carried out in more than one direction, including the use of new algorithms, hybrid algorithms, and the development of systems that enable fast and easy development of competent AI models in the banking arena. Robust explanations of AI-based models through model interrogation can contribute to unique insights and transparency, including causality.

The Gaussian process is the preferred choice of model because it is interpretable with a probabilistic interpretation. Extrapolating further, the basis of selection for analyses could be traced to the accurate predictive performance of the neural network. However, as the architecture becomes more complex, the complexity associated with interpretable weights and biases proliferates, and the potential inference becomes intractable. The Euclidean version of the DNN is taken into account for the analyses in this study, but given that other bases, including discrete data, will be created, it is worth mentioning that this methodology could be withdrawn. The second thrust of future work could pertain either to model interrogation or straightforward neural networks as they enhance their explanatory powers.

7. Conclusion

The in-depth analysis of AI/ML capabilities in different functional areas of business of a universal bank has been carried out. The emphasis has been on direct and indirect customer behavior analysis. Motivation is to maximize the capabilities of the most valuable bank's asset, i.e., modern digital IT. Oligopoly in the banking industry can become a kind of extreme competition, where only the best become winners and the weakest are thrown out of the market. It is believed that the materials reveal numerous applications that can be of interest and directly beneficial for various categories of readers, from students and teachers to operational bank teams and top executives. The tandem of modern digital technology and advanced AI capabilities are able to add value through the more sophisticated big data analysis and to use the obtained customer behavior insights for the sake of customer centricity.

The core idea suggests developing and supporting an end-to-end digital business ecosystem with digital bank that acts as the main driver in use of advanced AI methods simultaneously in all core functional areas leading to the operationalization of the existing and future innovative digital bank's digital business model. There are several guidelines concerning the most promising research areas for further evolution giving modern turbulent challenges.

Reference:

1. Tamanampudi, Venkata Mohit. "NLP-Powered ChatOps: Automating DevOps Collaboration Using Natural Language Processing for Real-Time Incident Resolution." Journal of Artificial Intelligence Research and Applications 1.1 (2021): 530-567.
2. Sangaraju, Varun Varma, and Kathleen Hargiss. "Zero trust security and multifactor authentication in fog computing environment." Available at SSRN 4472055.
3. S. Kumari, "Kanban and Agile for AI-Powered Product Management in Cloud-Native Platforms: Improving Workflow Efficiency Through Machine Learning-Driven Decision Support Systems", Distrib Learn Broad Appl Sci Res, vol. 5, pp. 867–885, Aug. 2019
4. Pal, Dheeraj Kumar Dukhiram, et al. "Implementing TOGAF for Large-Scale Healthcare Systems Integration." Internet of Things and Edge Computing Journal 2.1 (2022): 55-102.
5. Zhu, Yue, and Johnathan Crowell. "Systematic Review of Advancing Machine Learning Through Cross-Domain Analysis of Unlabeled Data." Journal of Science & Technology 4.1 (2023): 136-155.
6. J. Singh, "The Future of Autonomous Driving: Vision-Based Systems vs. LiDAR and the Benefits of Combining Both for Fully Autonomous Vehicles ", J. of Artificial Int. Research and App., vol. 1, no. 2, pp. 333–376, Jul. 2021
7. Gadhiraju, Asha. "Improving Hemodialysis Quality at DaVita: Leveraging Predictive Analytics and Real-Time Monitoring to Reduce Complications and Personalize Patient Care." Journal of AI in Healthcare and Medicine 1.1 (2021): 77-116.

8. Gadhiraju, Asha. "Empowering Dialysis Care: AI-Driven Decision Support Systems for Personalized Treatment Plans and Improved Patient Outcomes." Journal of Machine Learning for Healthcare Decision Support 2.1 (2022): 309-350.
9. Tamanampudi, Venkata Mohit. "Automating CI/CD Pipelines with Machine Learning Algorithms: Optimizing Build and Deployment Processes in DevOps Ecosystems." Distributed Learning and Broad Applications in Scientific Research 5 (2019): 810-849.
10. J. Singh, "Understanding Retrieval-Augmented Generation (RAG) Models in AI: A Deep Dive into the Fusion of Neural Networks and External Databases for Enhanced AI Performance", J. of Art. Int. Research, vol. 2, no. 2, pp. 258–275, Jul. 2022
11. S. Kumari, "Cloud Transformation and Cybersecurity: Using AI for Securing Data Migration and Optimizing Cloud Operations in Agile Environments", J. Sci. Tech., vol. 1, no. 1, pp. 791–808, Oct. 2020.
12. Sangaraju, Varun Varma, and Senthilkumar Rajagopal. "Applications of Computational Models in OCD." In Nutrition and Obsessive-Compulsive Disorder, pp. 26-35. CRC Press.
13. Tamanampudi, Venkata Mohit. "A Data-Driven Approach to Incident Management: Enhancing DevOps Operations with Machine Learning-Based Root Cause Analysis." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 419-466.