UNCOVERING USER PERCEPTIONS TOWARD DIGITAL BANKS IN INDONESIA: A NAÏVE BAYES SENTIMENT ANALYSIS OF TWITTER DATA

MULYANI KARMAGATRI1*, CLARISA FEZIA AMANDA AZIZ2, WINI RIZKI PURNAMA ASIH3, ISMA ADDI JUMRI4

1,2,3 Entrepreneurship Department. Bina Nusantara University. Jakarta, Indonesia
4Faculty of Technology Management & Technopreneurship, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
E-mail: 1 mulyani@binus.edu, 2 Clarisa.aziz@binus.ac.id, 3 wini.asih@binus.ac.id, 4 isma@utem.edu.my

Corresponding author: mulyani@binus.edu

ABSTRACT

The use of digital banks in Indonesia has rapidly increased in recent years in response to the adoption of new technologies and changes in consumer behavior. User responses to digital banks vary depending on their experience throughout their transactions on the application, which may result in satisfaction or dissatisfaction. Social media platforms such as Twitter have become a space for companies to obtain textual data related to customer reviews and their brand image. In this study, data obtained from Twitter have undergone the stages of data crawling and data cleaning. The subsequent stages involved classification using the Naïve Bayes algorithm and word cloud visualization to identify the most commonly used words based on user responses. The results of this study indicate that users' positive sentiment towards digital banks is influenced by the application's ease of use, while dissatisfaction is caused by technical constraints experienced during the administrative process. The positive, negative, and neutral sentiments in this study are used to identify business opportunities for digital banks and practical implications for future digital banking services.

Keywords: Digital Bank, Sentiment Analysis, Naïve Bayes, User Experience, Machine Learning

1. INTRODUCTION

The wave of new technologies has affected the financial industry, including changes in the traditional banking structure. Increasingly, innovations are being introduced in this sector as a form of meeting consumer demand. Recently, banking services have been developing special digital bank applications (also known as virtual or internet banks). The current dominance of digital banking is due to the increasing use of online and mobile banking platforms, as well as online and mobile payment solutions among smartphone and internet users [1]. Digital banks are the digitalization of banking transactions, including payment, deposit, transfer, and cash withdrawal utilizing the internet. Digital banks improve the customer experience by offering convenience [2]. Several advantages offered by digital banks include higher customer priority, virtual validation for account verification, more practical processes, lower administrative costs, minimal direct contact, and investment-related services that were previously limited to mobile banking and e-banking. Despite the added value generated by the digitalization process, combining greater interactivity of internet network systems in the banking mechanism has created a number of risks for users. The absence of a physical financial institution is the main factor causing user doubts. So far, there is no legal framework governing digital banks in Indonesia, so user data and financial protection are not yet guaranteed by the state. At the same time, users are also facing concerns about privacy and cyber security systems in these applications. The openness of the financial market due to deregulation and technology has replaced face-to-face contact, thereby increasing the likelihood of problems [3]. Digital banks face competition from non-bank companies and fintech, so banks must maintain existing customers while attracting new ones [4].

The advent of new technological waves has influenced the financial industry, including changes in traditional banking structures. The increasing demand from consumers has resulted in numerous innovations in this sector. In recent times, the
banking industry has been developing specialized digital bank applications (also known as virtual banks or internet banks). The current proliferation of digital banking dominance is due to the growing use of online and mobile banking platforms, as well as online and mobile payment solutions among smartphone and internet users [1].

Digital banks are the digitization of banking transactions, including payments, deposits, transfers, and cash withdrawals using the internet. Digital banks enhance the customer experience by offering convenience [2]. Some of the benefits offered by digital banks include higher priority for customers, virtual validation for account verification, more practical processes, lower administrative costs, minimal direct contact, and investment-related services that were previously limited to mobile banking and e-banking.

Despite the added value generated by the digitalization process, combining greater interactivity from internet network systems into banking mechanisms has created several risks for users. The absence of physical financial institutions is the main factor causing user uncertainty. So far, there is no legal framework governing digital banks in Indonesia, so user data protection and financial security are not guaranteed by the state. At the same time, users are also concerned about privacy and cyber security systems in these applications. Financial market openness resulting from deregulation and technology has replaced face-to-face contact, making the likelihood of problems greater [3]. Digital banks face competition from non-bank companies and fintech, so banks must retain existing customers while attracting new ones [4].

The realization of digital banks evokes different sentiments for each user, depending on the structural, procedural, accommodation, and digital bank product aspects that depend on individual experiences throughout the transaction process. To specifically determine the form of response, one platform that can be used as a data source is the social media platform Twitter. Text mining techniques such as sentiment analysis using Naïve Bayes classification algorithms are used to extract useful information and find data patterns from Twitter related to digital bank user experiences.

Analyzing the advantages and disadvantages of digital banks currently has significant managerial implications, as developers can identify features that need to be improved and prioritized. By understanding user opinions, businesses can see business opportunities and target better digital bank strategies to meet customer expectations. This study is designed to determine the parameters of the advantages and disadvantages of current digital banks based on Twitter user tweets in Indonesia, with a sub-focus on searching for the most frequently appearing keywords. This keyword search is general, meaning that this study does not search for the advantages and disadvantages of only one digital bank. These keywords can be an indication of the main advantages and complaints felt by digital bank users in Indonesia.

2. RELATED WORK

There are several indicators that affect user satisfaction, as well as dissatisfaction, with the use of digital banks. In order to achieve overall customer satisfaction, digital banks need to pay attention to several components in their products, such as cost, usability, trust, social influence, credibility, service responsiveness, and information privacy [5]. The numerous components that need to be considered are due to the different technology and business models adopted by digital banks compared to conventional banks. Data security is a major concern for customers, and online transactions need to be protected and guaranteed to be safe. In the digital environment, security is necessary to maintain authentication and authorization mechanisms between users, merchants, and service providers [5]. Improving the quality and qualifications of these components has a significant impact on customer satisfaction.

User responses to digital banks are expressed in the form of user reviews on various media. However, companies have difficulty tracking all reviews due to their unstructured and highly varied nature [6]. Furthermore, Martens [7] states in his research that users' emotional sentiment is weakly correlated with app ratings, proving that ratings are not an accurate representation of customer sentiment.

Several researchers have conducted studies on customer responses to digital banks using sentiment analysis techniques from text mining to delve deeper into responses. Cheng and Sharmayne [7] used LDA (Latent Dirichlet Allocation) topic modeling in their research to obtain relevant topics from digital banks and extract keywords from text based on frequency counts that will be labeled in digital banking features. The review obtained is divided based on positive and negative and association analysis is applied to determine the relationship between bank features and ratings given. The results show that some customers are interested in the early adoption of digital banking, especially with the fact that digital banks dare to offer higher interest rates than conventional banks. Flexibility towards service access is also a point of praise. However, the data
verification process, service quality, business integrity, deposit, transfer, and payment features need to be improved.

Another study was also conducted by Shetu [8] who predicted user satisfaction and dissatisfaction with online banking in Bangladesh using a machine learning approach using several algorithms including Naïve Bayes and Decision Tree while comparing the performance of all algorithms. In his study, the accuracy rate of the Naïve Bayes algorithm was 89%. The recall rate and f1 score were 85.3% and 94% respectively. As a result, the researcher found that about 51% of people were satisfied with online banking (which had easy access to online banking) and about 49% were dissatisfied. Although the results are fairly balanced, this data shows that there are still many aspects that need to be improved in online banking.

This research aims to uncover customer perceptions of online banking in Indonesia using a machine learning approach, specifically the Naïve Bayes algorithm. The researchers mined customer perceptions from Twitter social media. Hopefully, the results of this research can be used by stakeholders and online banking service providers to improve and maintain their services, and thus retain and grow their customer base.

3. THEORITICAL BACKGROUND

3.1. Digital Bank

The concept of "digital banking" has rapidly developed in Europe during the period from 2015 to 2020 [9]. Digital banking is a modern innovation that provides more personalized banking services to customers through mobile applications or personal computers. Digital banking enables customers to obtain information, communicate, and carry out banking transactions through the internet.

Digital banking differs from traditional banks in that it does not have physical branches, offers a wider range of services (such as more flexible savings and investment options), and provides virtual debit and credit cards. Cash transactions can actually disappear, and payments can be centralized on digital platforms. As a result of this innovation, digital bank customers can access banking services at any time with sophisticated IT support and high levels of flexibility. Customers do not need to go to branch offices to obtain services, thus facilitating faster and easier transaction processes.

Generally, digital banks do not expand their branch networks due to the 100% online communication system between the bank and the customer [9]. For digital banking companies, this service can reduce operating costs and optimize revenue based on costs when compared to transaction services through branch offices with high cost requirements. It is known that digital banks are divided into two types: those that have licenses and operate independently with their own brand, and those that collaborate with traditional banks and become subsidiaries of the parent bank [9].

3.2. Service Quality

The quality of service represents the gap between customer expectations of what a company should provide and the perceived performance of the service [10]. Other academics define service quality as the ability to adapt to client demands in the delivery of services [10]. Furthermore, the level of service quality affects post-purchase behavior and future individual decisions. Service quality also influences the level of customer loyalty to a product. Customer perceptions of quality are assumed to occur at various levels in service provision. Initially, customers evaluate the quality of interaction with service providers at an individual level, which is then extended to higher dimensional levels [4]. According to Parasuraman in DAM [10], the five dimensions of service quality can be described as follows: (1) Reliability, which denotes the service provider's capacity to deliver services consistently and accurately. (2) Tangibles, which encompasses the physical elements used to convey the service, such as the appearance of the physical facilities, equipment, and personnel. (3) Empathy, which pertains to the service provider's ability to understand and address the customers' needs and concerns. (4) Responsiveness, which denotes the service provider's willingness to provide prompt and efficient service to customers. (5) Assurance, which refers to the service provider's ability to inspire confidence in the customers by offering warranties, guarantees, and assurances of the quality of the service.

3.3 Customers Satisfaction

Consumers can indicate an estimate that determines how satisfied they are with a product, organization, or administration [11]. According to Oliver [12], customer satisfaction is defined as the post-consumption evaluation of a product or service, with a pleasant level of overall fulfillment related to product use. Fulfilling consumer needs with a sense of pleasure is considered a key driver of customer satisfaction in the context of product influence, equity, expectations, and/or performance. The level of customer satisfaction can also be demonstrated
through word-of-mouth reviews, customer behavioral intentions, and loyalty towards a product.

Furthermore, Oliver [12] also explains the timeframe of customer satisfaction, which is categorized into temporal and aggregate satisfaction. Temporal satisfaction is temporary and obtained from a single purchase, which has the potential to change due to other experiential variables in the future, while aggregate satisfaction describes satisfaction obtained from the accumulation of multiple purchase experiences.

Customer satisfaction can also be grouped based on the timing of evaluation, namely real-time evaluation (which takes place before and throughout the purchasing process) and post-purchase evaluation (evaluation obtained based on the customer's experience using the product). Post-purchase evaluation will have a greater impact on future decisions because it is rooted in a stronger end result. Customers like these tend to repurchase from the same company [13].

The performance of a company in serving customers is also positively correlated with the level of customer satisfaction. Price in [14] defines performance as synonymous with organizational performance effectiveness and identifies productivity, suitability, and institutionalism as criteria for recognition. Moh [14] proposes another view that productivity, flexibility, and adaptability of the organization are indicators of company performance. Customer loyalty data, including reviews and ratings, can assist organizations in deciding the best ways to improve quality or modify their products and administration [11].

Customer service is crucial for any company, including those in the digital banking industry. In the early stages of a product, companies need to monitor user experience and customer expectations. Satisfied customers are more likely to remain loyal. This research aims to fill a gap by providing authentic customer perceptions of digital banking. By mining customer tweets on Twitter, the researchers can gain insight into their feelings, experiences, and expectations.

4. METHOD

In this study, several steps were taken to achieve the desired results. These steps followed the Cross-Industry Standard Process for Data Mining (CRISP-DM), which is a standardized process for managing data mining projects. Data mining techniques used within the CRISP-DM process may vary, including sentiment analysis. In the context of digital banking in Indonesia, sentiment analysis can be used to understand public perception and opinion trends regarding the services and products offered by digital banks. This can help digital banks understand customer satisfaction and improve their services, as well as identify potential business opportunities for the digital banking industry. CRISP-DM consists of six steps, as follows.

A. Business Understanding

This study aims to analyze the sentiments and define the satisfaction of the public, specifically the users of digital banks in Indonesia. The objectives of this research are to investigate "What are the sentiments of digital bank users in Indonesia?", "How is the sentiment analysis of user satisfaction with digital banks in Indonesia conducted using Naïve Bayes?", and "What are the business opportunities and practical implications that can be derived from the results of this research?".

B. Data Understanding

In this study, the dataset was obtained from the Kaggle website titled "Sentiment Analysis of Bank digitals in Indonesia". The dataset consists of a collection of Twitter user opinions regarding satisfaction with various digital bank services in Indonesia. The tweets were collected from October 2016 to February 2022, mentioning digital banks such as Jago, Digibank, Livin by Mandiri, Neocommerce, Jenius, Raya, TMRW, Blu, Sea Bank, Line Bank, and Wokee.

C. Data Preparation

The collected data has undergone data crawling and data cleaning to eliminate retweets and duplicates. Next, the data was manually labeled as positive, negative, or neutral. The data preparation procedure was then continued to the data transformation stage. At this stage, the cases were transformed to lower case to ensure uniformity, and the data was further filtered using Indonesian stop words. To reduce the data for easier management and relevance, the data underwent tokenization.

D. Modeling

The methodological algorithm employed in this research is Naïve Bayes. The labeled data will be trained with Naïve Bayes algorithm. Naïve Bayes is chosen because it often produces better prediction results than other classification techniques[17].
5. RESULT AND DISCUSSION

Data processing was carried out using Rapid Miner data processing tools. The dataset used in this research consisted of Twitter user opinions related to the keyword "digital bank". Following the research objectives to determine Twitter user opinions on digital banks in Indonesia, all data in the dataset were in the Indonesian language. The total data in the dataset was 2390, and it had been cleaned from retweets and duplications.

The obtained dataset had gone through data cleaning and data crawling processes. Next, the data will be manually labeled as shown in Table X, with positive labels such as "great," "helpful," "easy," "cool," and "satisfactory." Negative labels include "disappointed," "slow," "please," "improve," and "failed."

When numbering equations, enclose numbers in parentheses and place flush with right-hand margin of the column. Equations must be typed, not inserted. (If nonstandard fonts are used its better to put equations as images instead of text)

The next stage involves data transformation, stopword filtering, and data reduction through tokenization. Once the data is prepared, the next step is to create a model. The data will be trained using the Naive Bayes algorithm model.

<table>
<thead>
<tr>
<th>Text</th>
<th>Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Great app pelayanan memuaskan untuk app banking tidak banyak kendala dan mudah digunakan. (Great app, the service is satisfactory for a banking app with minimal issues and easy to use)</td>
<td>Positive</td>
</tr>
<tr>
<td>Lemot aplikasi lemot bikin lama untuk bertransaksi (the app is slow and causes delays in transactions)</td>
<td>Negative</td>
</tr>
<tr>
<td>Setelah di perbaharui mudah han lebih baik lagi (After being updated, the app is now easier to use and has improved significantly)</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

The performance of the applied model is measured using the performance operator. The following are the results of the testing using the Naive Bayes model.

The next stage involves data transformation, stopword filtering, and data reduction through tokenization. Once the data is prepared, the next step is to create a model. The data will be trained using the Naive Bayes algorithm model.
Based on the research conducted, it was found that using the Naïve Bayes model, the accuracy rate obtained was 88.12%. This means that the sentiment analysis model using Naïve Bayes is proven to be good. In the sentiment analysis, the number of Twitter user tweets about digital banks in Indonesia that were considered positive amounted to 1330 tweets, but machine learning provided results that only 1108 tweets were considered truly positive with a recall percentage of 83.31%. Meanwhile, for the number of tweets considered negative, there were 656 tweets, but machine learning provided results that only 594 tweets were truly negative with a recall percentage of 90.55%. As for the number of neutral tweets, both before and after using machine learning, they remained the same at 404. Thus, neutral sentiment tweets have a recall rate of 100%. From these results, it can also be concluded that the distribution of sentiment that is most prominent in customer satisfaction with digital banking services in Indonesia is positive sentiment. The distribution of positive sentiment is greater than neutral and negative sentiment.

Using word cloud, the most dominant words can be found, as presented in figure 4. Word cloud is a visual representation of text data that visualizes the frequency of words used in the data. Word cloud is very useful in sentiment analysis because it helps identify the most frequently used words in the data and shows trends and patterns in word usage.

In the analysis of sentiment towards digital banks in Indonesia, word cloud can be used to identify the most frequently used words in user reviews or social media posts about digital banks. These words can then be aligned with the data available to be classified based on sentiment (positive, negative, and neutral) and analyzed to obtain information about what customers are looking for or questioning about digital banks. Thus, word cloud helps provide a visual representation that is easy to read from text related to digital bank sentiment, making it easier to determine the strengths and weaknesses of digital banking services in Indonesia for improvement and future business opportunities.

By aligning these words with the content of user tweets about digital banking services, the detailed meaning of these words can be known. The alignment results are explained in table 5.

<table>
<thead>
<tr>
<th>Table 4 : Keywords description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word</strong></td>
</tr>
<tr>
<td>Mudah (easy)</td>
</tr>
<tr>
<td>Membantu (helpful)</td>
</tr>
<tr>
<td>Keren (cool)</td>
</tr>
<tr>
<td>Ribet (complicated)</td>
</tr>
<tr>
<td>Login</td>
</tr>
<tr>
<td>Verifikasi (verification)</td>
</tr>
<tr>
<td>Update</td>
</tr>
<tr>
<td>Cepat (quick)</td>
</tr>
<tr>
<td>Mantap (Excellent)</td>
</tr>
</tbody>
</table>
Based on the sentiment analysis and the description of the most frequently used words, a businessperson can identify potential business opportunities in digital banking. This study provides information on the advantages of doing business in digital banks, such as the fact that banking activities are predominantly conducted online and users highly prefer this mode of banking. Consequently, the costs associated with establishing a digital bank are lower than those of traditional banks with more offline activities, including the costs of creating cards and hiring employees for physical service locations.

However, digital banks have their weaknesses because they require continuous updating, which can result in complaints from users. These updates must be conducted regularly to ensure that the provided applications improve and are in line with the latest developments. Furthermore, when building a digital bank, the application's features should be as comprehensive as possible. As the sentiment analysis indicates, the features of digital banks are highly valued by users. For example, if a digital bank does not support online payments for a particular product or platform, such as social media or e-commerce, users will switch to another digital bank that provides complete payment features. Another weakness that can be found in digital banking businesses is the need for a stable and reliable infrastructure so that customers will not experience difficulties accessing the service, which is the majority of complaints in this analysis.

Bank digital presents a huge business opportunity in the current landscape. This opportunity can be attributed to the fact that digital banking has become a necessity, where payment activities are shifting towards online platforms, such as e-commerce. Additionally, the image of digital banks is generally perceived as positive by the public as it simplifies and expedites transaction processes. E-commerce and social media shopping platforms, which are now directly connected to digital banks, are some of the opportunities for building a digital bank business.

However, the threat of building a digital bank lies in the banking activities that are often impractical and prone to failure due to poor system performance. Errors and system failures can hinder and damage overall banking activities. Furthermore, poor system performance can lead to other threats, such as negative user ratings. In the word cloud analysis, the word "stars," which is related to ratings, dominated. This means that when users give poor ratings to digital banking applications, they may share their ratings with the public on social media. A negative image in the public eye can deter people from using the digital banking application.

The results of this study can also provide recommendations for the development of digital banking services in the future to maintain the positive aspects of digital banking, such as ease of use, comprehensive features, and fast processes. Conversely, the development of digital banking services in the future should avoid and address negative aspects of digital banking services identified in the sentiment analysis, such as difficulty in the login process, verification failures, and complicated registration procedures.

Digital banking has a significant opportunity today, as more people use technology and the internet for their financial activities. From user tweets, it is evident that digital banking offers convenience, speed, and security in transactions, which leads to many people shifting from traditional ways of conducting financial transactions. There are many digital banking businesses, including payment applications, e-wallets, and financial technology services. The opportunity for digital banking businesses will continue to grow and evolve with the growth of technology and the internet.

### 6. CONCLUSION

Based on the sentiment analysis of Twitter data on digital banking in Indonesia using the Naïve Bayes algorithm, the following conclusions can be drawn: The accuracy rate of sentiment analysis using Naïve Bayes was evaluated as good with an accuracy percentage of 88.12%. In the sentiment analysis, the number of tweets from Twitter users regarding digital banking in Indonesia that were considered positive was 1330, but machine learning produced results indicating that only 1108 tweets were actually positive with a recall rate of 83.31%. Additionally,
for the number of tweets considered negative, which was 656, machine learning produced results indicating that only 594 tweets were actually negative with a recall rate of 90.55%. The number of neutral tweets, both before and after using machine learning, remained the same, totaling 404 tweets. The recall rate for neutral sentiment was 100%. Thus, positive sentiment was found to be the most dominant sentiment in this analysis.

In this analysis, the positive sentiment of users towards digital banks in Indonesia is generally based on the ease of using the application, the helpfulness of digital banking services in transactions, the cool and complete features available in them, and the quick process in conducting transfers. On the other hand, the negative sentiment of users towards digital banks in Indonesia is generally based on the difficulty of the registration process, difficult login procedures, and frequent failures during verification. Digital banking presents a good business opportunity because it is currently a necessity and users prefer the speed and convenience of online transactions. However, in building a digital bank, regular updates and a good system are necessary.

Customer service is crucial for any company, including those in the digital banking industry. In the early stages of a product, companies need to monitor user experience and customer expectations. Satisfied customers are more likely to remain loyal. This research contributes to filling a gap by providing authentic customer perceptions of digital banking. By mining customer tweets on Twitter, the researchers can gain insight into their feelings, experiences, and expectations. Practical implications for future digital banking services include maintaining the positive aspects mentioned in the positive sentiment and removing and improving the negative aspects mentioned in the negative sentiment. The results of this research can be used by stakeholders and online banking service providers to improve and maintain their services, and thus retain and grow their customer base.

REFERENCES:


