Uncovering Key Service Improvement Areas in Digital Finance: A Topic Modeling Approach Using LDA on User Reviews

Jalel Ben Othman[1,*], Taqwa Hariguna[2]

[1]University of Paris 13, France

[2]Magister of Computer Sciences, Amikom Purwokerto University, Indonesia

ABSTRACT

The rapid expansion of digital finance has transformed the way financial services are accessed and utilized, particularly in emerging markets such as Indonesia. This study aims to uncover key service improvement areas within the Easycash mobile lending platform by analyzing user reviews through topic modeling using Latent Dirichlet Allocation (LDA). The research employed a data-driven approach, combining text preprocessing in Bahasa Indonesia using the Sastrawi library, TF-IDF vectorization, and sentiment classification with machine learning models including Naive Bayes, K-Nearest Neighbors (KNN), and XGBoost. The XGBoost model achieved the highest performance with an F1-score of 0.9709, effectively distinguishing between positive, neutral, and negative sentiments. LDA analysis identified five major topics: Loan Limits and Repayment, Customer Gratitude and Satisfaction, Loan Application Process and Interest Rates, App Quality and Customer Service, and Data Management and Account Issues. Results indicate that while Easycash users generally express positive sentiment toward ease of use and service speed, concerns persist regarding high interest rates, customer service responsiveness, and data privacy. These findings provide actionable insights for fintech companies to enhance user satisfaction through targeted service improvements and continuous feedback analysis.

Keywords Digital Finance, Fintech, User Reviews, Sentiment Analysis, Topic Modeling, Latent Dirichlet Allocation (LDA), Machine Learning, Customer Satisfaction, Service Improvement

Introduction

The rapid growth of financial technology (fintech) has reshaped the global digital finance landscape, expanding access to financial services for diverse populations. Advanced technologies such as artificial intelligence (AI), blockchain, and big data have been at the forefront of this transformation, enabling more efficient and accessible financial systems. These innovations have had a profound impact, especially in developing regions, where they have been shown to improve total factor productivity and mitigate economic stagnation. The integration of fintech into the broader digital economy has democratized access to financial services, allowing both banked and unbanked individuals to participate in the financial system. This is particularly important for marginalized groups, who have historically been excluded from traditional financial institutions [1], [2]. As mobile financial services become a mainstay for conducting transactions, especially during the COVID-19 pandemic, fintech

Submitted 5 June 2025
Accepted 6 August 2025
Published 13 November 2025

Corresponding author
Jalel Ben Othman, jalel.ben-othman@univ-paris13.fr

Additional Information and Declarations can be found on page 456

DOI: 10.47738/jdmdc.v2i4.47

How to cite this article: J.B. Othman, T. Hariguna, "Uncovering Key Service Improvement Areas in Digital Finance: A Topic Modeling Approach Using LDA on User Reviews," *J. Digit. Mark. Digit. Curr.*, vol. 2, no. 4, pp. 434-460, 2024.

companies must remain responsive to user feedback to enhance their services continually.

Fintech has acted as a key driver of financial inclusion, broadening the reach of financial services through digital platforms. This growing reliance on mobile applications and digital finance solutions underscores the critical role of user feedback in shaping the direction of fintech innovations. Continuous improvement in these services depends on how effectively fintech companies gather, analyze, and respond to user needs. As fintech companies grow and their services evolve, the importance of user experience becomes even more central to maintaining trust and fostering loyalty among their users [2], [3]. The iterative process of integrating feedback allows fintech firms to not only enhance user satisfaction but also adapt more effectively to the competitive pressures exerted by traditional financial institutions, driving innovation across the financial sector [4], [5]. Thus, understanding user perspectives and addressing service-related issues becomes crucial in the fast-paced digital finance ecosystem.

In Indonesia, fintech companies such as Easycash have transformed the traditional loan application process by offering fast, unsecured loans via mobile applications. This shift reflects a broader trend in the Indonesian financial landscape, where the demand for digital finance solutions has grown rapidly. The widespread adoption of mobile technology in Indonesia, where over 90% of the population uses mobile phones, has enabled fintech firms to reach a broad audience, particularly those underserved by traditional banking systems [6]. These advancements have allowed users to access credit services with unprecedented ease and speed, enhancing financial inclusion and convenience for millions of Indonesians.

The success of these fintech platforms is heavily dependent on the quality of user experience they provide. Research has demonstrated that user satisfaction with fintech applications is largely driven by the perceived ease of use, efficiency, and reliability of the service [7], [8]. User-friendly interfaces and secure transaction processes are essential in building trust with consumers, especially in a digital environment where security concerns are paramount. Fintech companies like Easycash have thus focused on optimizing their services and improving customer service to ensure a seamless experience. Such improvements are crucial for retaining users and promoting growth in an increasingly competitive fintech market [9].

In addition to the focus on user experience, the regulatory framework governing fintech in Indonesia plays a pivotal role in the sector's development. Authorities such as the Financial Services Authority (OJK) and Bank Indonesia oversee fintech operations, ensuring that companies like Easycash comply with financial regulations while fostering innovation [10]. However, as the fintech sector expands, there remains a need for adaptive regulations to address the challenges posed by unregulated or illegal lending practices [11], [12]. As the regulatory landscape evolves, the success of fintech platforms will increasingly depend on their ability to balance innovation with compliance, ensuring both user trust and financial stability.

User satisfaction played a critical role in determining the success of fintech companies like Easycash. In the competitive landscape of digital finance, companies were required to consistently monitor and improve their services to meet user expectations and retain their customer base. User feedback, especially in the form of reviews, provided valuable insights into user experiences and satisfaction. However, given the large volume of user reviews generated on platforms like Easycash, manually analyzing this feedback could have been more efficient and practical. Fintech companies needed to adopt more advanced, automated methods to process this data and extract actionable insights efficiently.

The challenge of dealing with vast amounts of unstructured data, such as textual reviews, posed a significant barrier for Easycash in leveraging this valuable resource for service improvement. Traditional manual review processes were not only time-consuming but also prone to human error and inconsistency. As a result, important trends or issues could easily go unnoticed, hindering the company's ability to make timely adjustments or improvements based on user feedback. Moreover, without a systematic approach to categorizing and understanding the diverse topics discussed in user reviews, it became difficult to prioritize areas that required immediate attention.

Given these limitations, there was a pressing need for fintech companies to implement automated techniques, such as topic modeling, to analyze user reviews at scale. These techniques allowed companies like Easycash to uncover recurring themes or issues from thousands of reviews without the need for manual intervention. By identifying key service areas that needed improvement, companies could respond more efficiently to user concerns, thereby enhancing overall customer satisfaction and fostering loyalty in a highly competitive market. The application of tools like Latent Dirichlet Allocation (LDA) offered a structured and scalable approach to processing user feedback, which would prove essential in driving continuous service improvement in the digital finance sector.

While the use of sentiment analysis in the fintech sector has been widely explored, most studies primarily focused on analyzing user sentiments—whether positive, negative, or neutral. Sentiment analysis provided valuable insights into the overall mood of user feedback. Still, it often fell short of offering detailed, topic-specific information that could pinpoint the precise areas where improvements were needed. For example, a user could leave a negative review. Still, without further analysis, it would be unclear whether the dissatisfaction stemmed from poor customer service, high interest rates, or technical issues within the app. As fintech companies continued to grow, the need for more granular insights into user feedback became essential.

Recent research in sentiment analysis and machine learning has revealed valuable insights into consumer behavior, e-commerce trends, and financial transactions across digital platforms. Studies on aspect-based sentiment analysis have shown that using semantic similarity can enhance the classification of visitor reviews, improving the ability to capture nuanced consumer sentiments and preferences [13]. Additionally, research on discount strategies and consumer ratings has highlighted how promotional tactics can significantly influence customer perceptions, demonstrating the importance of tailored sentiment analysis approaches to accurately interpret consumer feedback in digital marketplaces [14]. The adoption of clustering algorithms for customer segmentation, such as K-Means and DBSCAN, has proven effective in categorizing users based on behavior, enabling more personalized marketing strategies in e-commerce [15]. Similarly, predictive modeling techniques, including decision trees and random forests, have been applied to estimate campaign ROI, offering a data-driven approach to strategic decision-making in digital marketing [16].

In the realm of financial analysis, studies have investigated the relationship between trading volume and Bitcoin price movements using correlation methods, providing insights into market dynamics within digital currencies [17]. Moreover, unsupervised anomaly detection techniques in cryptocurrency trading have highlighted the utility of clustering and density-based approaches for identifying irregularities in transaction patterns [18]. Expanding into the Web3 space, research has explored the blockchain and cryptocurrency job market, revealing emerging trends and employment opportunities within this sector [19]. Furthermore, studies on virtual property markets have examined factors influencing sales trends and price determinants, showcasing the role of blockchain technology in shaping digital real estate [20]. These studies collectively underscore the diverse applications of machine learning and data analysis techniques across digital finance, marketing, and blockchain-based sectors, emphasizing their potential to uncover patterns and inform strategic decisions in an increasingly digital economy.

Latent Dirichlet Allocation (LDA) offered a solution to this gap by allowing researchers and companies to move beyond the broad categories provided by sentiment analysis and dive deeper into the specific topics or themes present in large datasets of textual feedback. Although LDA has been successfully implemented in various industries to analyze large volumes of text, its application in the Indonesian fintech context, particularly with platforms like Easycash, has been limited. This lack of research highlighted a significant gap, as fintech companies in Indonesia were expanding rapidly, and their success heavily depended on continuously improving user experience. A more detailed, topic-based analysis could provide actionable insights into the specific service areas users frequently mentioned in their reviews.

In the context of Indonesian digital finance, applying LDA to user reviews represented a relatively unexplored avenue for enhancing service quality. Most existing research on fintech feedback analysis in Indonesia primarily employed traditional sentiment analysis methods, which, while useful, needed more depth to identify recurring service-related themes. Addressing this research gap, the study aimed to apply LDA to analyze Easycash user reviews, thereby uncovering key topics that could inform service improvement strategies. This novel approach provided an opportunity to contribute both to the academic literature on fintech in Indonesia and the practical application of advanced text analysis methods in a rapidly growing industry.

This research aimed to uncover key service improvement areas in the Easycash mobile application by analyzing user reviews. Easycash, like many fintech platforms, relied on user feedback to identify potential areas of service enhancement. However, given the large volume of reviews available, manually extracting insights from this unstructured data posed a significant challenge. Therefore, this study sought to implement a data-driven approach to automate the process of discovering recurring themes and topics within the user feedback. The objective was to provide actionable insights into the specific aspects of the service that required attention, ultimately improving the overall user experience.

To achieve this goal, the study employed topic modeling using LDA, an unsupervised machine learning algorithm well-suited for uncovering hidden topics in large text datasets. LDA allowed for the identification of distinct themes within user reviews by analyzing the co-occurrence patterns of words across multiple documents. In this case, the documents were individual reviews, and the analysis was focused on the content field of the dataset, which contained the textual feedback from Easycash users. By applying LDA, the study aimed to move beyond broad sentiment analysis and provide a more granular understanding of user concerns and preferences.

The specific focus of the research was on identifying service areas frequently mentioned in the reviews, such as customer support, loan approval processes, or interest rates. These insights were expected to help Easycash prioritize improvements in areas that directly impacted user satisfaction. Furthermore, this approach provided a scalable solution for continuously monitoring user feedback, enabling Easycash to respond more efficiently to emerging issues and maintain a competitive edge in the digital finance landscape.

Literature Review

Digital Finance and User Experience in Fintech

User reviews have become an integral component of the digital finance landscape, playing a significant role in influencing user engagement, product development, and overall market dynamics. As digital financial services continued to grow, user-generated content—particularly in the form of reviews—proved to be essential in shaping consumer perceptions and behaviors. These reviews not only provided direct insights into the experiences of users but also acted as a form of social proof, which enhanced trust and credibility in financial products and services. The ability to assess the opinions and feedback of other users allowed potential customers to make informed decisions, reducing the perceived risks associated with adopting digital financial solutions.

The incorporation of user reviews into digital finance platforms demonstrated several benefits, particularly in enhancing user engagement and satisfaction. Reviews often provide personalized feedback, which companies could use to tailor financial products to the specific needs of their users, thus improving the overall user experience. For example, personalized recommendations based on feedback helped reduce information search costs and increased user retention rates [21]. Additionally, user reviews helped to address the issue of information asymmetry that was common in financial markets, especially for small and medium-sized enterprises (SMEs) seeking financing. By fostering a more transparent environment where users could share their experiences, digital finance platforms encouraged informed decision-making among potential customers, thereby improving the credibility of the platforms themselves.

Moreover, user reviews had a direct impact on the ongoing development and refinement of financial products. By collecting feedback from users, fintech companies were able to identify areas that needed improvement and innovate accordingly. For instance, reviews that highlighted issues related to the usability of an app or concerns about security could prompt the development of more user-friendly interfaces or the introduction of new security. This iterative process of product development, driven by user feedback, was crucial in maintaining competitiveness in the fast-paced digital finance sector. Furthermore, the influence of user reviews extended beyond individual products to shape broader market trends. Positive reviews could enhance the reputation of a fintech platform, thereby attracting new users, while negative reviews, if not addressed promptly, could lead to a decline in trust and engagement [2].

Customer feedback was essential for improving service quality in FinTech platforms, as it directly influenced user satisfaction, retention, and the overall development of financial services. In the competitive fintech landscape, where user expectations continued to rise, understanding and integrating customer feedback into service offerings provided a competitive advantage. Feedback mechanisms, such as reviews and surveys, allowed fintech companies to assess the quality of their services and identify areas where improvements were necessary. This process was particularly important in building user trust, as companies that responded to user concerns were more likely to retain their customer base [22]. Studies showed that fintech platforms that actively solicited and analyzed feedback had better customer satisfaction, which directly impacted user loyalty and engagement [23].

The importance of customer feedback was further highlighted in guiding the refinement of fintech services. Users often provided insights into their experiences, pinpointing specific aspects of the service that required improvement, such as ease of navigation, transaction speed, or the clarity of information provided by the platform. These insights were invaluable for fintech companies seeking to enhance their services, as they enabled the development of more user-friendly and efficient systems. In turn, this customer-centric approach contributed to a stronger user experience, which was crucial for the adoption of fintech services. Research demonstrated that when users found services easy to use and reliable, their likelihood of continued use and recommendation increased significantly [24].

Moreover, customer feedback played a pivotal role in fostering trust between fintech platforms and their users. Trust was a critical factor in digital finance, where users relied on the security and transparency of the services provided. Fintech companies that acted on customer feedback demonstrated their commitment to improving the user experience, thus fostering a sense of reliability and security [25]. Additionally, the ongoing process of collecting and analyzing user feedback enabled fintech companies to stay ahead of market trends and evolving user needs. This proactive approach helped companies maintain their relevance in a rapidly changing industry and supported their innovation efforts to meet user expectations and demands [26].

Related studies on analyzing customer reviews in fintech services, particularly within Southeast Asia, have focused on the role of customer feedback in shaping user experience and driving service improvements. The rapid growth of fintech platforms in the region, especially in Indonesia and Malaysia, has created a strong need for companies to understand their users' preferences and experiences better. Research [27] emphasized that customer experience is central to fintech success, as innovations must blend traditional financial services with modern technologies to meet evolving user expectations. Their research highlighted that customer feedback serves as a valuable resource for assessing how well these innovations are being received, thereby guiding further improvements.

Research [28] explored the specific relationship between user innovativeness and fintech adoption in Indonesia, concluding that understanding user perspectives through feedback is essential for the continuous development of fintech services. They argued that by analyzing customer reviews, fintech companies could better tailor their products to meet user needs, thus fostering higher adoption rates. This study aligns with the broader literature on customer feedback in Southeast Asia, underscoring its importance in driving product enhancements that resonate with the local user base.

In a broader regional context, research [29] conducted a systematic literature review that demonstrated the impact of consumer behavior on fintech adoption, particularly focusing on perceived benefits such as convenience and security. The review indicated that customer reviews reflect these perceived benefits, offering fintech companies critical insights for product development and marketing. Similarly, study by [30] in Malaysia applied the Technology Acceptance Model (TAM) to investigate consumer attitudes toward fintech products. Their findings suggested that perceived usefulness and ease of use, often derived from customer feedback, were key factors influencing adoption, further emphasizing the role of reviews in shaping fintech services in the region.

Topic Modeling in Natural Language Processing (NLP)

LDA was widely recognized as a powerful statistical model in the field of Natural Language Processing (NLP), primarily used for topic modeling. The model operated under the assumption that documents were a mixture of various topics, with each topic being characterized by a distribution of specific words. This generative approach allowed researchers to uncover hidden thematic structures in large collections of textual data. LDA's ability to detect latent topics made it particularly valuable for analyzing unstructured text where predefined categories were unavailable. Its application was demonstrated across a variety of fields, such as healthcare, finance, and social media, where it proved to be effective in extracting insightful information from vast amounts of data [31].

LDA served as a foundational tool in topic modeling, offering the ability to identify, categorize, and cluster themes within extensive text datasets. This technique proved especially useful in fields requiring large-scale data analysis, such as clinical decision-making, where understanding underlying topics in medical literature could aid in the diagnosis and treatment of patients [32], [33]. Additionally, LDA was applied in analyzing healthcare data, including patient safety concerns, which helped extract relevant insights from various textual sources [34]. The model's capacity to delineate key topics made it instrumental in many industries, as it not only facilitated keyword identification but also provided meaningful groupings of related documents, enhancing the clarity of large datasets [35].

The unsupervised learning nature of LDA was one of its strongest features, allowing it to explore textual data without the need for predefined labels. This flexibility made it ideal for exploratory data analysis, particularly in domains where the data structure was not fully understood. In fields such as genomics, LDA was used to reveal mutation signatures and classify genomes, while in social media analysis, it helped uncover themes from user-generated content, shedding light on public sentiment [36], [37]. The model's hierarchical Bayesian framework enabled it to handle complex relationships between topics and documents, making it an invaluable tool for large-scale data analysis, and its impact extended into public health and financial sectors where policy decisions could be informed by LDA-driven insights [38], [39].

Latent Dirichlet Allocation (LDA) is a probabilistic generative model commonly used in Natural Language Processing (NLP) for topic modeling. It assumes that each document in a corpus is a mixture of topics, and a distribution of words characterizes each topic. The primary goal of LDA is to discover the hidden thematic structure in large collections of text. The model calculates the probability of a word being part of a topic given its document context as shown in the formula:

$$P(w|z) = \sum_{d=1}^{D} P(w|z,d)P(z|d) \tag{1}$$

This framework allows LDA to identify latent topics by analyzing the co-occurrence patterns of words across documents.

Text Preprocessing for the Indonesian Language

Processing Indonesian text in the context of Natural Language Processing (NLP) presented unique challenges, particularly due to the morphological complexity of the language. One of the primary techniques used to address this complexity was stemming, which reduced words to their root forms to facilitate more accurate analysis. The Sastrawi library, a tool specifically designed for the Indonesian language, played a critical role in this process. Sastrawi implemented the Nazief-Adriani stemming algorithm, which has proven effective in handling various morphological forms of Indonesian words, including affixes and reduplications [40], [41]. This made it an essential tool for text preprocessing in applications like topic modeling, where accurate word representation was crucial.

In addition to stemming, the Sastrawi library also provided functionalities for stopword removal, which was another critical step in text preprocessing. Stopwords, or frequently used words that added little value to text analysis, needed to be filtered out to improve the efficiency of NLP models. The Sastrawi library included an extensive list of common Indonesian stopwords, enhancing the quality of the processed text before further analysis. Studies demonstrated that using Sastrawi for both stemming and stopword removal significantly improved the accuracy of text-based models, such as in sentiment analysis, by reducing noise and standardizing text input [42], [43]. These preprocessing steps were vital in transforming raw text into a format that NLP algorithms could effectively analyze.

However, challenges in the stemming process, particularly with over-stemming and under-stemming, remained persistent. Over-stemming occurred when unrelated words were incorrectly reduced to the same root while under-stemming resulted when related words were not sufficiently reduced to their base form. These issues could skew the results of text analysis, impacting the performance of the overall model. Despite Sastrawi's strong performance, research indicated that improvements were still needed to reduce such errors, especially in comparison to other stemming algorithms like the Paice-Husk method [44], [45]. Moreover, the prevalence of slang and informal expressions, particularly in social media and user-generated content, further complicated text preprocessing. To address this, researchers suggested expanding the Sastrawi stopword list to include commonly used slang terms, ensuring a more robust handling of real-world text [42].

The removal of stopwords and the handling of colloquial language, including slang, are essential steps in processing Indonesian language text, especially for tasks such as sentiment analysis and topic modeling. Stopwords are common words like conjunctions and prepositions that do not carry significant meaning in text analysis. Their removal improves the efficiency of Natural Language Processing (NLP) tasks by reducing noise in the data. This leads to better accuracy in models used for indexing, classification, and information retrieval [46], [47]. In Indonesian text processing, the Sastrawi library is frequently utilized for this purpose, providing a robust mechanism for eliminating unnecessary words and improving the clarity of text analysis [42].

In addition to stopwords, Indonesian slang presents another challenge in text preprocessing, especially in user-generated content such as reviews or social media posts. Slang refers to informal language variations that often differ from standard vocabulary and are not always recognized by conventional NLP tools. This can lead to inaccuracies or misinterpretations in sentiment analysis. Studies have demonstrated that transforming slang into its formal equivalent, as defined in the Indonesian Dictionary (KBBI), enhances model performance, with improvements in accuracy observed in various tasks [48]. For instance, converting colloquial expressions to their standard forms has been shown to increase the accuracy of sentiment analysis models by up to 3.5% [49], highlighting the importance of adapting preprocessing methods to account for the dynamic nature of informal language.

The frequent use of slang, particularly among younger populations in Indonesia, further underscores the importance of incorporating this linguistic variation into text preprocessing. Studies have shown that nearly 66.7% of Indonesian students regularly use slang in daily communication, signaling a significant cultural shift towards informal language usage in both social and digital contexts [50], [51]. As slang continues to evolve, often influenced by regional dialects and foreign languages, NLP systems must continually update their slang dictionaries and preprocessing algorithms to maintain relevance [52]. By addressing these variations in language, NLP models can capture the true meaning behind user-generated content, ultimately reducing the effectiveness of tasks like sentiment analysis and text classification [53]. Therefore, integrating stopword removal and slang handling into preprocessing pipelines is crucial for accurate and contextually relevant text analysis.

Exploratory Data Analysis (EDA) for Text Data

EDA is essential for understanding the distribution and patterns present in textual data. It acts as a preliminary but critical step in the data analysis process, enabling researchers to gain insight into the overall structure and quality of their data before applying more advanced techniques. EDA employs methods such as visualizations, word frequency analysis, and statistical summaries to reveal trends, outliers, and anomalies. This initial exploration helps researchers identify key patterns that may influence the outcome of subsequent analyses, ensuring that the data is well-prepared for further processing [54].

One of the primary functions of EDA is to assess the quality of the text data, especially when dealing with datasets sourced from user reviews, social media, or other unstructured formats. Textual data often contains inconsistencies, missing values, and noise, which can hinder the performance of machine learning models if not addressed. Through EDA, researchers can clean the data by identifying and removing irrelevant elements, such as extraneous punctuation or redundant terms, which ensures that the final dataset is representative and reliable [55]. Techniques like tokenization and lemmatization, commonly employed during EDA, further refine the text by breaking it down into manageable units, making it easier to uncover significant patterns in word usage and thematic distribution.

Additionally, EDA plays a vital role in preparing text data for machine learning tasks, such as text classification and topic modeling. Through frequency analysis, EDA helps uncover the most common words and phrases, revealing underlying themes or sentiments within the dataset [56]. This exploratory phase also supports the application of data augmentation techniques, particularly in scenarios where the dataset is small or imbalanced. As [57] demonstrated, methods like synonym replacement or random insertion can enrich the dataset, improving the model's ability to generalize to new data. The insights gained from EDA guide researchers in optimizing their models by ensuring the text data is both comprehensive and well-structured, providing a strong foundation for more complex analytical procedures.

Visualization techniques play a crucial role in the EDA of textual data, offering clear and intuitive ways to interpret complex information. These techniques, such as word clouds, bar charts, and term frequency distributions, help researchers uncover patterns and trends that might take time to emerge from raw text. Each method provides distinct insights, making them invaluable tools for gaining an initial understanding of the data before moving on to more advanced analysis methods. Their ability to quickly summarize and visually represent textual information is essential in extracting key insights from large datasets.

Word clouds, one of the most recognizable visualization tools, display the most frequent words in a dataset by adjusting their size based on frequency. Larger words indicate higher frequency, while smaller words represent less common terms. This technique offers an immediate visual snapshot of the dominant themes within a dataset, making it a popular starting point for textual analysis. The simplicity and effectiveness of word clouds allow researchers to identify important words and themes quickly, which is particularly useful when exploring unfamiliar datasets. Despite their aesthetic appeal, word clouds do not provide precise quantitative information, but they are highly effective in giving a broad overview of textual content [58].

Bar charts and term frequency distributions offer a more detailed and precise way to visualize textual data. Bar charts allow for clear comparisons between the frequencies of different terms, providing a more structured and quantifiable approach than word clouds. Researchers can use bar charts to easily identify the most and least frequent terms within a dataset, making it easier to detect patterns or anomalies. Similarly, term frequency distributions offer a statistical view of word usage across a dataset, typically represented through histograms. These distributions can highlight important features such as skewness or outliers, providing deeper insights into the structure of the text. Both techniques are essential in revealing trends, supporting the identification of key topics, and preparing the data for more complex analyses like sentiment analysis or topic modeling [59].

Method

The research method for this study consists of several steps to ensure a comprehensive and accurate analysis. The flowchart in figure 1 outlines the detailed steps of the research method.

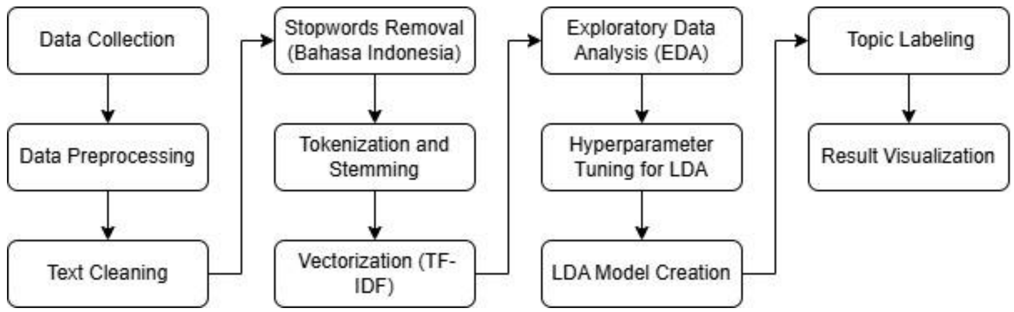

Figure 1 **Research Method Flowchart**

Data Description

The dataset analyzed in this research comprises user reviews collected from the Easycash platform, a financial technology application operating in Indonesia. The reviews primarily focus on the `content` column, which contains the textual feedback provided by users, offering insight into their experiences and satisfaction with the service. Additional columns in the dataset include `score`, which records user ratings on a numeric scale, and `at`, which logs the timestamp of each review. These additional columns support the analysis by providing quantitative indicators of user sentiment and a temporal context for each review.

To facilitate a comprehensive understanding of user feedback on Easycash, the dataset undergoes an initial descriptive analysis. First, the total number of reviews is calculated, providing a quantitative foundation for subsequent analyses. This overview reveals the scope of user engagement and feedback available for analysis, helping to contextualize the significance of identified patterns. Following this, the average score is determined, offering a preliminary insight into overall user satisfaction. The mean score serves as a benchmark, indicating the general sentiment trend across all reviews and setting a baseline for deeper sentiment analysis and topic exploration.

The dataset is further examined through a temporal analysis of the review distribution, which is conducted using the `at` column. This analysis reveals how user engagement with the Easycash platform has evolved over time and highlights any potential seasonal trends or periods of heightened activity. For example, increases in review volume following significant app updates or marketing campaigns may indicate times when user feedback is particularly critical. Visualizing the review distribution over time aids in identifying these patterns, helping to connect fluctuations in user engagement with potential changes in user sentiment or emerging service improvement areas.

Data Preprocessing

The data preprocessing phase began with text cleaning to ensure consistency and readability in the user review content. This process involved the removal of punctuation, numbers, and special characters, all of which are extraneous in natural language processing (NLP) tasks and can introduce noise into the analysis. The next step was stopword removal, a crucial step in minimizing irrelevant words. Given that this study focused on reviews in Bahasa Indonesia, the Sastrawi library—a tool specifically designed for processing Indonesian text—was utilized to eliminate common stopwords in the language. This step reduced the dataset's word count to only the most meaningful terms, enhancing the clarity of insights derived from the data.

Following stopword removal, the text was tokenized, breaking down each review into individual words. Tokenization enables detailed text analysis by facilitating word-by-word examination. After tokenization, the Sastrawi library was again employed, this time for stemming. Stemming converts words to their root forms, which is particularly useful for Indonesian, a language with extensive word variations due to prefixes, suffixes, and infixes. Stemming allows words with the same root meaning, such as "membantu" and "bantuan," to be treated as a single term, thereby reducing dimensionality and improving the efficiency of subsequent analysis steps.

The final stage of data preprocessing involved converting the cleaned and stemmed text into a numerical form using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. TF-IDF is an effective method for representing text data as it assigns weight to words based on their frequency within a document relative to their frequency across all documents. By emphasizing terms that are more unique to individual reviews, TF-IDF helps capture the distinctive language used by users when expressing specific concerns or praises. For this study, the TF-IDF vectorizer was configured with a `max_features` parameter to limit the vocabulary size and ensure model performance optimization. This matrix served as the input for the LDA model in the topic modeling phase, enabling the identification of key themes in user feedback.

Sentiment Analysis

The sentiment analysis phase of this study focused on categorizing user reviews into positive, negative, or neutral sentiments to provide insights into user satisfaction with Easycash services. Initial sentiment labeling was conducted using an automated sentiment analysis tool, with options such as VADER or a pre-trained sentiment model. In cases where sentiment scores were not directly available, labels were derived based on thresholds applied to the `score` column, with higher scores indicating positive sentiment and lower scores indicating negative sentiment. This process allowed for the consistent classification of sentiments across the dataset, setting a foundation for further analysis.

Following the labeling process, three machine learning models—Naive Bayes, K-Nearest Neighbors (KNN), and XGBoost—were employed to classify sentiment. Each model was trained on the labeled dataset to evaluate its effectiveness in distinguishing between sentiment categories. Naive Bayes, commonly used in text classification due to its simplicity and efficiency, served as a baseline model. KNN, which classifies data points based on their proximity to neighbors, was included to assess its performance on this text-based task. XGBoost, a powerful ensemble technique known for its accuracy in predictive tasks, rounded out the comparison. These models were selected for their varying approaches to classification, enabling a comprehensive evaluation of model suitability for sentiment analysis in this context.

To ensure reliable results, each model's performance was validated through k-fold cross-validation, with k = 5, which allowed for an assessment across multiple data subsets. This method provided average scores for metrics including accuracy, precision, recall, and F1-score, offering a balanced view of each model's strengths. Based on these metrics, the model that achieved the highest overall performance was selected as the most effective for sentiment classification. The results from this evaluation phase provided insights into which techniques were most suitable for understanding user sentiment in the Easycash reviews, forming the basis for further topic modeling and detailed sentiment-based analysis.

Exploratory Data Analysis (EDA)

The EDA phase began with the creation of a word cloud to visualize the most frequently occurring words in the Easycash user reviews, shown in figure 2.

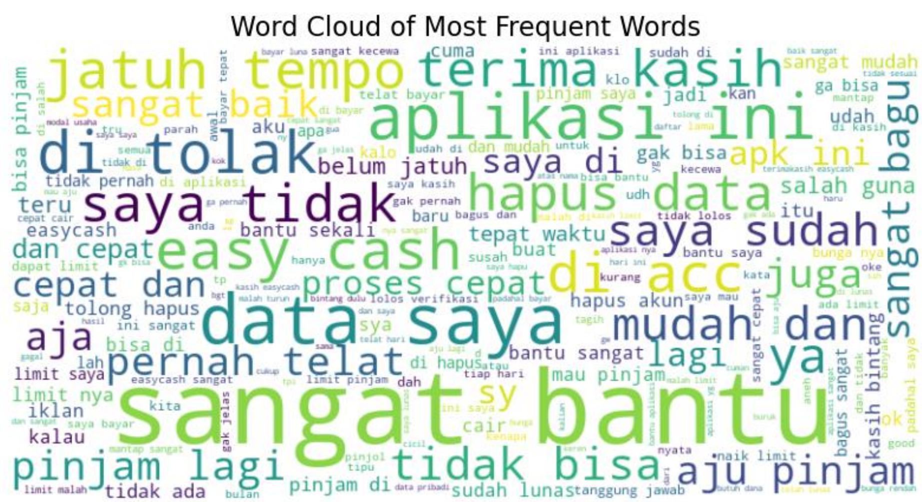

Figure 2 **Word Cloud of Most Frequent Words**

The word cloud generated from Easycash user reviews highlights key themes and recurring words, providing a visual summary of user sentiments and concerns. Prominent words such as "sangat," "bantu," and "cepat" suggest that many users appreciate the helpfulness and speed of the service, indicating positive experiences with the application's efficiency in processing loans. The words "mudah" and "proses" further reinforce this theme, suggesting that users find the loan application process straightforward. This positive feedback is tempered by other words, like "tolak" and "jatuh tempo," which imply some dissatisfaction, particularly around issues of loan rejections and deadlines.

Additionally, the frequent mention of "data," "saya," and "hapus" suggests that users may have concerns about data management and privacy, or issues with account deletion. Terms like "tempo," "bunga," and "lunas" reflect user focus on repayment terms, interest rates, and loan clearance, indicating that financial terms are a significant aspect of the user experience. Phrases such as "terima kasih" (thank you) and "sangat baik" (very good) also appear, signaling gratitude and satisfaction from certain users. Conversely, words like "tidak" and "lagi" could imply frustrations, as they often appear in contexts where users express what they are unable to do or what they choose not to repeat. Overall, this word cloud reveals a mix of positive experiences with ease of use and service speed, balanced with concerns around loan rejections, data handling, and repayment terms. To further quantify the importance of specific terms, a bar plot of term frequencies was generated, shown in figure 3.

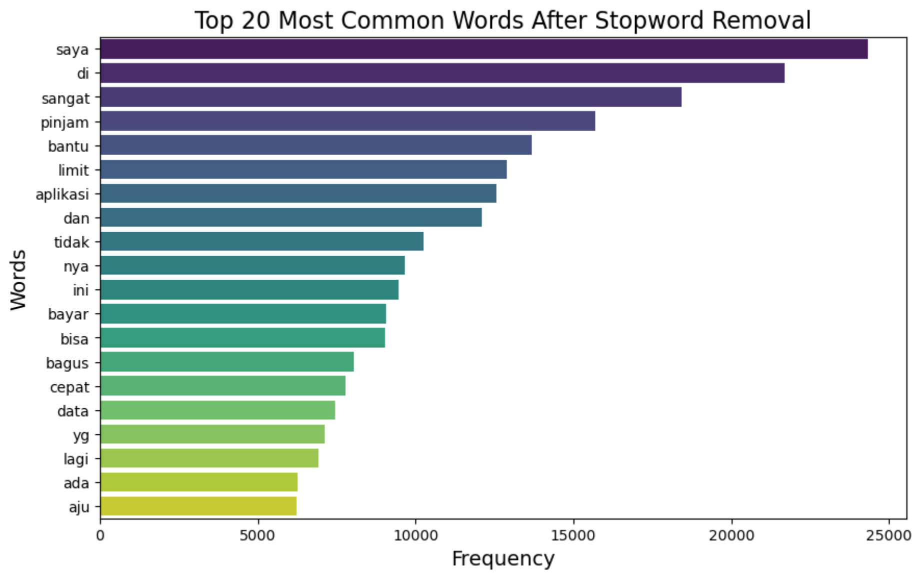

Figure 3 Most Common Words After Stopword Removal

The bar chart displays the 20 most common words in Easycash user reviews after removing stopwords, revealing key themes and user concerns. The words “saya” (I) and “di” (at/in) are among the most frequent, suggesting that users often reference their personal experiences or specific contexts within their feedback. High frequencies of words like “sangat” (very), “bantu” (help), and “bagus” (good) indicate common expressions of satisfaction, highlighting positive aspects of the service. This also implies that users frequently emphasize the degree of their experiences, whether positive or negative.

On the other hand, words such as “pinjam” (borrow), “limit,” “tidak” (not), and “bayar” (pay) point to transactional aspects of the application that users discuss most often. The presence of “limit” and “tidak” suggests concerns or issues related to borrowing limits or limitations in the service, while “bayar” emphasizes payment-related discussions. The prevalence of terms related to transactions, assistance, and functionality underscores the practical concerns users express in their reviews, indicating areas for potential improvement, such as lending limits and application usability. This analysis offers insights into the primary focus areas of user feedback and can guide future service enhancements.

By combining the word cloud and bar plot, the EDA phase provided a solid foundation for understanding user priorities and recurring topics in the reviews. These visualizations offered valuable context for interpreting the results of subsequent topic modeling and sentiment analysis. Specifically, they highlighted core themes that might align with different sentiment categories and topics, helping to inform the analytical steps that followed in identifying key service improvement areas for Easycash.

Topic Modeling using LDA

The topic modeling phase used Latent Dirichlet Allocation (LDA) to uncover key themes within the Easycash user reviews. To determine the optimal number of topics, hyperparameter tuning was conducted by testing various topic counts and evaluating their coherence scores. This step ensured that the chosen number of topics offered a meaningful balance between interpretability and comprehensiveness. Coherence scores served as an indicator of topic quality, reflecting how well the identified terms within each topic aligned semantically. By maximizing coherence scores, the model captured well-defined topics that minimized redundancy and enhanced interpretability. Algorithm 1 outlines the process of topic modeling using Latent Dirichlet Allocation (LDA) to uncover dominant themes within user reviews. This algorithm involves determining the optimal number of topics through coherence score evaluation, training the final LDA model on the TF-IDF matrix, and assigning each review to its most representative topic. The pseudocode also details the extraction, labeling, and interpretation of topics, providing a structured overview of how recurring themes were identified and analyzed from the user feedback dataset.

Algorithm 1 Topic Modeling Using Latent Dirichlet Allocation (LDA)

Data Preparation

• Load the preprocessed dataset $D = \{r_1, r_2, \dots, r_N\}$, where each $r_i$represents a cleaned and tokenized user review.
• Build the vocabulary $V = \{w_1, w_2, \dots, w_m\}$consisting of all unique words from D.
• Generate a TF-IDF matrix $X \in \mathbb{R}^{N\times m}$, where each element $x_{i,j}$denotes the TF-IDF weight of word $w_j$in document r_i.

Hyperparameter Tuning for Optimal Topic Count

• Define a set of candidate topic numbers $K = \{5,10,15,20,25\}$.
• For each $k \in K$:
  – Initialize an LDA model LDA$_k$with $k$topics.
  – Fit the model: LDA$_k$ ← fit(X).
  – Compute the topic coherence score $C_k$using:
    $C_k = \frac{1}{|T_k|}\sum_{t\in T_k}$ Coherence(t),
    where $T_k$represents the set of topics for model LDA$_k$, and Coherence(t)measures semantic similarity among top words within topic t.
• Select the optimal number of topics $k^* = \arg\max_{k\in K}(C_k)$, corresponding to the model with the highest coherence score.

Model Training with Optimal Parameters

• Initialize the final LDA model: LDA* = LDA$(k = k^*)$.
• Train the model using the TF-IDF matrix: LDA* ← fit(X).
• For each document r_i:
  – Compute the topic probability distribution:
    $\theta_i = [p(t_1 \mid r_i), p(t_2 \mid r_i), \dots, p(t_{k^*} \mid r_i)]$.

– Assign the primary topic to the document:
$t_i = \arg\max_{t_j} p(t_j \mid r_i)$.

• Store the document-topic mapping $D_T = \{(r_i, t_i)\}_{i=1}^{N}$.

Topic Extraction and Representation

• For each topic $t_j \in \{1,2, \dots, k^*\}$:
  – Retrieve the top $n$most representative words:
  $W_j = \text{TopWords}(t_j, n) = \{w_{j1}, w_{j2}, \dots, w_{jn}\}$.
  – Calculate the topic-word distribution:
  $\phi_{j,w} = p(w \mid t_j)$ for all $w \in W_j$.

• Output the complete list of topics:
$T^* = \{(t_j, W_j, \phi_j)\}_{j=1}^{k^*}$.

Topic Labeling and Interpretation

• For each topic $t_j \in T^*$:
  – Review the top associated words W_j.
  – Assign a human-readable label $L_j$based on semantic relevance.
  Examples:
  • If $W_j$contains {"interest", "payment", "fee"} → $L_j =$ "Interest and Payment Issues".
  • If $W_j$contains {"service", "response", "support"} → $L_j =$ "Customer Service".
  • If $W_j$contains {"easy", "fast", "simple"} → $L_j =$ "Ease of Use".

• Save the labeled topics as $T_L = \{(t_j, L_j, W_j)\}_{j=1}^{k^*}$.

Evaluation and Visualization

• Visualize coherence scores across topic counts by plotting $C_k$versus $k$to confirm the optimal k^*.
• Generate word clouds or bar plots for each topic $t_j$to illustrate the most frequent and influential words.
• Calculate topic proportions across all documents to identify the most prevalent themes.
• Summarize findings by highlighting key labeled topics $L_j$that reflect recurring user feedback patterns.

End.

After selecting the optimal number of topics, the LDA model was applied to the TF-IDF matrix created from the preprocessed text data. Each review in the dataset was assigned a primary topic based on the model's output, which represented the most prominent theme associated with that review. This process transformed the textual data into a set of distinct topics, each representing a collection of terms frequently mentioned together in user feedback. The LDA model allowed for the identification of underlying patterns in the reviews, providing a structured understanding of recurring themes within the user feedback on Easycash services.

The final step involved topic labeling, where each topic was manually interpreted based on its most prominent terms. By examining the top words associated with each topic, meaningful labels were assigned to reflect the main themes in user concerns and experiences. For example, topics that included terms such as "bunga" (interest) and "bayar" (pay) were labeled as relating to interest rates and fees, while topics featuring words like "layanan" (service) and "respon" (response) were categorized under customer service. These labeled topics provided clear and interpretable insights into the areas of Easycash that users discussed most frequently, forming the basis for understanding key service improvement areas.

Result and Discussion

Sentiment Analysis Results

The sentiment analysis of Easycash user reviews revealed insights into the overall user satisfaction with the platform, categorized as positive, neutral, or negative. A temporal analysis of sentiment distribution (shown in figure 4) highlighted trends over time, with positive reviews consistently dominating the feedback throughout the year. Neutral sentiments followed, while negative sentiments remained significantly lower in comparison. This distribution suggests that the majority of users expressed a generally positive outlook on their experience with Easycash, although the presence of neutral and negative sentiments indicates room for improvement in certain areas.

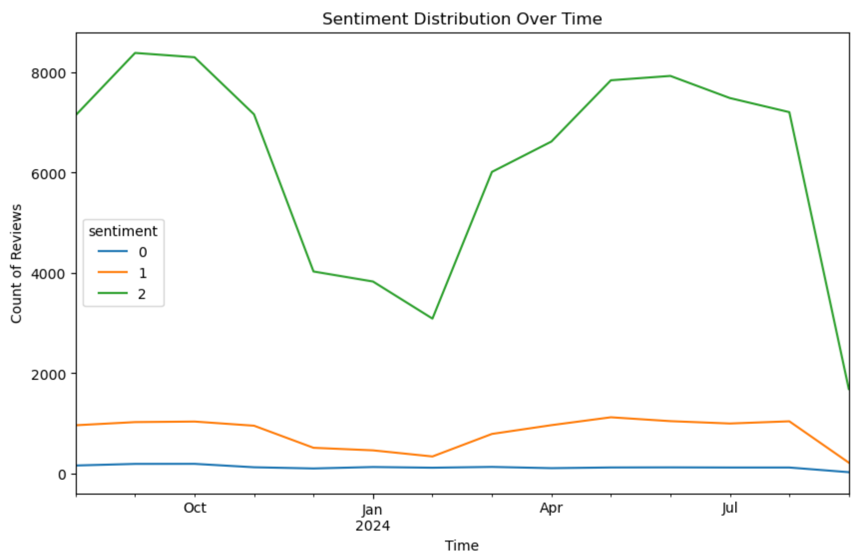

Figure 4 Sentiment Distribution Over Time

The line graph illustrates the distribution of review sentiments (positive, neutral, and negative) over time for Easycash user reviews. The green line, representing positive sentiment, dominates throughout the timeline, with a noticeable dip in early 2024, followed by a gradual increase until mid-year. This fluctuation in positive reviews might reflect changes in user satisfaction due to service updates, promotional activities, or seasonal trends affecting user experiences. The orange line indicates neutral sentiment, which remains relatively stable with minor variations over time. This stability suggests that a consistent portion of users perceived the service as adequate but not exceptional, likely influenced by factors that neither significantly enhanced nor detracted from their experience. Negative sentiments, represented by the blue line, remain consistently low, indicating a smaller subset of users expressing dissatisfaction. The steadiness of negative sentiment over time implies that significant service issues are not pervasive but may reflect isolated or individual user concerns. Overall, the sentiment distribution over time highlights an overall positive user experience with periodic fluctuations that could be explored further to identify potential service improvement opportunities.

To accurately classify user sentiments, three machine learning algorithms—Naive Bayes, K-Nearest Neighbors (KNN), and XGBoost—were implemented and evaluated using precision, recall, and F1-score metrics. Each model's performance was measured on the labeled Easycash dataset to determine its effectiveness in detecting sentiment polarity across reviews.

As summarized in table 1, XGBoost achieved the highest overall performance, with a precision of 0.9716, recall of 0.9720, and an F1-score of 0.9709. These results indicate XGBoost's strong ability to accurately identify sentiment classes with minimal misclassification. Naive Bayes and KNN also produced respectable results, recording F1-scores of 0.9295 and 0.9204, respectively, though both fell slightly short of XGBoost's precision and recall metrics.

Table 1 **Performance Comparison of Sentiment Classification Models**

Model	Precision	Recall	F1-Score	Interpretation
XGBoost	0.9716	0.9720	0.9709	Achieved the highest accuracy and balance across metrics; highly reliable for sentiment detection.
Naive Bayes	0.9334	0.9260	0.9295	Strong baseline performance; slightly lower precision compared to XGBoost.
K-Nearest Neighbors (KNN)	0.9240	0.9170	0.9204	Performs adequately but less consistent across sentiment categories.

Note. Metrics are averaged over k-fold cross-validation to ensure consistency and minimize evaluation bias.

The comparative results confirm XGBoost as the most effective classifier for sentiment analysis in this dataset. Its superior precision and recall values indicate that the model effectively captures linguistic nuances in user-generated text, particularly in multi-class sentiment settings. In contrast, Naive Bayes, though computationally efficient, demonstrated a slightly lower precision, reflecting occasional misclassifications in context-dependent reviews. KNN, while performing competitively, relies on distance-based similarity and may struggle with high-dimensional text representations, accounting for its marginally lower accuracy.

The high overall performance of all three models confirms that user sentiments within Easycash reviews can be reliably detected through supervised learning. However, XGBoost's gradient-boosting approach provides an additional edge by effectively minimizing classification errors and adapting to complex linguistic patterns present in user feedback. This superior performance underscores XGBoost's suitability for large-scale sentiment analysis tasks in financial technology applications, where precision and contextual understanding are essential for accurately interpreting user satisfaction and guiding service improvements.

Topic Modeling Results

The topic modeling analysis, conducted using Latent Dirichlet Allocation (LDA), identified five main topics within the Easycash user reviews. Each topic was labeled based on its most prominent terms, providing insights into recurring themes in user feedback. For instance, Topic 1, associated with terms such as "mantap" (great), "limit," "pinjam" (borrow), and "bayar" (pay), was labeled as Loan Limits and Repayment. This topic reflects users' positive remarks about borrowing limits and repayment flexibility. Topic 2, which includes words like "terimakasih" (thank you), "puas" (satisfied), and "keren" (cool), was labeled as Customer Gratitude and Satisfaction, indicating high satisfaction among users regarding Easycash's assistance during financial needs.

Additional topics provided further insights into the user experience. Topic 3, featuring terms such as "cepat" (fast), "mudah" (easy), and "bunga" (interest), highlighted aspects of the Loan Application Process and Interest Rates. This topic suggests that while users appreciate the ease of the loan application process, there may be mixed sentiments about the associated interest rates. Topic 4, with prominent terms like "bagus" (good), "respon" (response), and "aplikasi" (application), focused on App Quality and Customer Service. This topic pointed to positive user experiences with the app's responsiveness and quality, yet it also underscored areas for potential improvement in customer service interactions. Lastly, Topic 5 included terms such as "baik" (good), "data," and "hapus" (delete), and was labeled Data Management and Account Issues, suggesting that users encountered challenges with account-related issues and data handling. The analysis of topic distribution by sentiment, as visualized in figure 5, revealed how different topics correlated with positive, neutral, and negative sentiments.

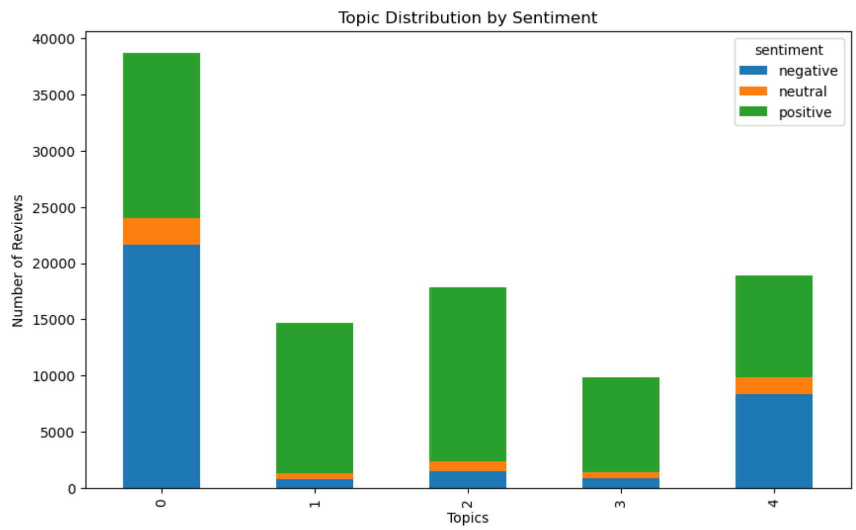

Figure 5 Topic Distribution by Sentiment

The bar chart above illustrates the distribution of review sentiments (positive, neutral, and negative) across five distinct topics identified from Easycash user reviews. Each topic is represented by a grouped bar, with colors indicating sentiment classifications. Topic 0 has the highest volume of reviews, encompassing a substantial portion of negative sentiment. This likely points to a prevalent issue affecting user satisfaction, such as high interest rates or application difficulties, which may have led to increased dissatisfaction. Positive sentiment is also present within Topic 0, suggesting some users still found favorable aspects even when critical issues were raised. Other topics, such as Topic 2 and Topic 4, show a more balanced sentiment distribution, with a prominent amount of positive feedback. This pattern indicates areas where Easycash is performing well, potentially reflecting efficient customer service or the app’s ease of use. Topics with a higher proportion of positive sentiment provide insights into strengths that could be emphasized or built upon to improve overall user satisfaction.

These results underscore the importance of each topic in shaping user perceptions of Easycash's services. Topics with a high proportion of positive sentiment, such as those reflecting satisfaction with customer service or loan accessibility, highlight aspects of the service that resonate well with users. Conversely, topics with significant neutral or negative sentiment, particularly those related to interest rates and data management, point to areas where Easycash could focus its service improvement efforts. The sentiment distribution across these topics provides a nuanced understanding of user experiences, revealing both strengths and areas for enhancement in Easycash's digital finance offerings.

Discussion

The topic modeling results offered a detailed view of user feedback on Easycash, revealing both positive and negative themes associated with the service. For instance, the Loan Limits and Repayment topic, which included terms like "mantap" (great) and "limit," demonstrated users' satisfaction with borrowing limits and the ease of repayment options. Reviews in this category generally praised Easycash for providing accessible loan amounts that align with users' financial needs. In contrast, the Interest Rates topic, which involved terms like "bunga" (interest) and "bayar" (pay), reflected a mix of sentiments. While some users appreciated the speed and convenience of the loan process, others expressed concerns about high interest rates, suggesting that fees associated with loans could be a point of dissatisfaction.

Further analysis of the Customer Service topic, which included terms such as "respon" (response) and "layanan" (service), highlighted mixed experiences with Easycash's customer support. Positive reviews commended the responsiveness and helpfulness of the support team, particularly during urgent financial situations. However, some users reported delays or unhelpful responses, indicating that improvements in customer service could enhance overall user satisfaction. This topic underscores the importance of maintaining efficient and empathetic customer support, as timely and effective communication significantly influences user perceptions of the service.

The Data Management and Account Issues topic, identified by terms like "data," "hapus" (delete), and "baik" (good), pointed to challenges related to account handling and data security. Reviews in this category often mentioned issues with account setup, data privacy, or the process of account deletion. While some users appreciated the app's security measures, others expressed concerns about data handling practices. This topic highlights the need for Easycash to reinforce its data security protocols and streamline account management processes to build trust and confidence among users.

In addition to interpreting the content of these topics, the study examined correlations between topics and average user scores to understand how specific issues influence overall satisfaction. Topics related to loan accessibility and quick processing generally correlated with higher scores, indicating that users value the convenience and efficiency of Easycash's loan services. Conversely, topics associated with high interest rates or customer service challenges often received lower scores, suggesting that these aspects may negatively impact user satisfaction. This correlation analysis helps to identify which areas are most crucial for Easycash to prioritize for improvement.

Lastly, a temporal analysis of topic trends revealed fluctuations in the prominence of certain issues over time (figure 6)

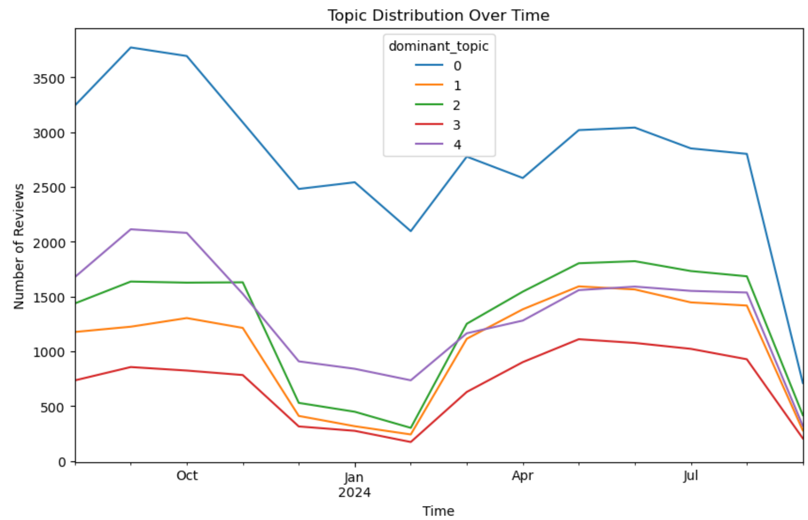

Figure 6 Topic Distribution over Time

The line graph above presents the distribution of topics over time, showing how each topic fluctuates across different periods. Notably, Topic 0 has a recurring presence, with a noticeable peak early on, suggesting persistent issues or concerns related to this topic that consistently attract user attention. The decline in mentions of Topic 0 towards the end of the timeline may indicate improvements in the areas it represents or a shift in user focus to other aspects of the service. Topics 2 and 4 show similar patterns, with a steady rise around April, which could correspond to specific events or updates that heightened user interest or satisfaction. The periodic dips and peaks across all topics reflect possible changes in service features, promotions, or seasonal influences impacting user experiences.

Overall, this temporal analysis highlights patterns in user feedback and can help Easycash pinpoint periods where particular topics are emphasized, thereby enabling targeted improvements during high-activity periods. Understanding these trends over time is crucial for identifying areas requiring continuous attention versus those that vary due to external factors or user demographics. These findings provide actionable insights into user expectations and satisfaction, highlighting areas where Easycash can make targeted improvements to enhance its digital finance services.

The analysis of topic distributions by sentiment, provides insights into the general sentiment associated with each topic identified through the LDA model. Topics such as Loan Limits and Repayment (Topic 0) are predominantly linked with positive sentiment, indicating users' satisfaction with borrowing limits and repayment terms. However, this topic also includes a significant portion of negative sentiment, suggesting that some users may experience issues despite the generally positive feedback. Other topics, like Interest Rates and Customer Service, show a more balanced distribution of sentiments, with notable proportions of neutral and negative feedback. This sentiment analysis by topic helps identify areas where Easycash can focus on enhancing user satisfaction.

The temporal analysis of topic trends, highlights how the prominence of specific topics fluctuates over time. For example, Customer Service issues appear to spike periodically, possibly reflecting increased user interaction during promotional events or app updates. Similarly, concerns about Interest Rates tend to show higher prominence at the start of financial quarters, indicating that users are particularly sensitive to fee structures during these periods. This temporal trend analysis is valuable for Easycash as it enables the company to anticipate periods when certain topics are more likely to surface in user feedback, allowing for proactive service improvements.

Additionally, the relationship between word count and sentiment, as illustrated in the third visualization, reveals interesting patterns in how review length correlates with expressed sentiment. Negative reviews tend to have a higher median word count, indicating that dissatisfied users often provide more detailed feedback. This suggests that negative experiences may prompt users to elaborate on their issues, potentially due to frustration or a desire to provide constructive criticism. In contrast, positive and neutral reviews are generally shorter, which may indicate a straightforward or concise expression of satisfaction or ambivalence. Understanding this relationship helps Easycash prioritize longer, detailed reviews in its analysis, as these reviews are likely to contain specific insights into user pain points.

This visualization-driven analysis highlights the significant insights gained by examining sentiment, temporal trends, and review length in conjunction with topic modeling results. By integrating these perspectives, Easycash can identify not only what users are discussing but also when and how these discussions evolve over time. For instance, identifying recurring customer service issues during peak usage times can prompt Easycash to allocate more resources to customer support during these periods, thereby potentially reducing negative feedback.

Conclusion

This study identified several key topics within Easycash user reviews using Latent Dirichlet Allocation (LDA) and assessed the associated sentiments to understand user satisfaction. The primary topics included Loan Limits and Repayment, Interest Rates, Customer Service, and Data Management and Account Issues. Sentiment analysis revealed that users generally expressed positive sentiments towards Easycash's loan accessibility and repayment options. However, topics related to high interest rates and customer service challenges often attracted neutral or negative sentiments, indicating that these areas may detract from overall user satisfaction. These findings highlight specific aspects of the service that influence user perceptions and satisfaction.

The insights from this analysis suggest actionable steps for Easycash to improve user satisfaction. For example, the prevalence of negative sentiments related to Interest Rates indicates that users may perceive the rates as high, which could discourage continued use of the service. Reducing interest rates or offering more transparent fee structures could enhance trust and improve user retention. Additionally, the analysis showed that Customer Service issues, such as delayed responses, impact user satisfaction. Strengthening the customer support system with faster response times and more effective issue resolution could significantly enhance the user experience. Simplifying the loan application process further would likely boost positive sentiment and attract new users by making the platform more user-friendly and accessible.

While this study provided valuable insights, it faced certain limitations. The LDA model, while effective in identifying topics, sometimes grouped disparate sentiments under the same topic, which made interpretation challenging. Additionally, the model may not fully capture the nuances of colloquial language, especially in Bahasa Indonesia, where slang and regional dialects vary widely. Sentiment analysis based solely on polarity (positive, neutral, negative) lacks the depth to capture subtler emotional tones, such as frustration or enthusiasm, which could provide more detailed insights into user experiences. Addressing these limitations could improve the accuracy and depth of future analyses.

Future research could build on this study by integrating more advanced sentiment analysis tools, such as emotion detection or aspect-based sentiment analysis, to capture a broader range of emotional responses. These tools could offer deeper insights into specific aspects of the service that elicit strong emotions, such as user trust or frustration. Additionally, applying these methods to user reviews from other fintech applications could provide a comparative perspective, enabling Easycash to benchmark its services against competitors. Such comparative studies would reveal broader trends in user expectations across the digital finance sector, guiding strategic improvements.

Declarations

Author Contributions

Author Contributions: Conceptualization, J.B.O. and T.H.; Methodology, J.B.O. and T.H.; Software, J.B.O. and T.H.; Validation, J.B.O. and T.H.; Formal Analysis, J.B.O.; Investigation, T.H.; Resources, J.B.O.; Data Curation, T.H.; Writing—Original Draft Preparation, J.B.O.; Writing—Review and Editing, T.H.; Visualization, T.H. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. K. Ozili and D. Mhlanga, “Why Is Financial Inclusion So Popular? An Analysis of Development Buzzwords,” *J. Int. Dev.*, vol. 36, no. 1, pp. 231–253, 2023, doi: 10.1002/jid.3812.

[2] P. K. Ozili, “Information Effect of Fintech and Digital Finance on Financial Inclusion During the COVID-19 Pandemic: Global Evidence,” *Fintech*, vol. 3, no. 1, pp. 66–82, 2024, doi: 10.3390/fintech3010005.

[3] C. C. Vergara and L. F. Agudo, “Fintech and Sustainability: Do They Affect Each Other?,” *Sustainability*, vol. 13, no. 13, p. 7012, 2021, doi: 10.3390/su13137012.

[4] I. Anagnostopoulos, “Fintech and Regtech: Impact on Regulators and Banks,” *J. Econ. Bus.*, vol. 100, pp. 7–25, 2018, doi: 10.1016/j.jeconbus.2018.07.003.

[5] I. M. Unal and A. F. Aysan, “Fintech, Digitalization, and Blockchain in Islamic Finance: Retrospective Investigation,” *Fintech*, vol. 1, no. 4, pp. 388–398, 2022, doi: 10.3390/fintech1040029.

[6] S. Narayan, “Does Fintech Matter for Indonesia’s Economic Growth?,” *Bull. Monet. Econ. Bank.*, vol. 22, no. 4, pp. 437–456, 2019, doi: 10.21098/bemp.v22i4.1237.

[7] R. Apriliani, “The Influence of Intention to Use Digital Wallet Applications, E-Service Quality and Trust on Consumer Satisfaction Toward Digital Payment Applications,” *J. Inf. Dan Teknol.*, pp. 125–131, 2024, doi: 10.60083/jidt.v6i1.485.

[8] M. Y. Edward, E. N. Fu’ad, H. Ismanto, A. D. R. Atahau, and R. Robiyanto, “Success Factors for Peer-to-Peer Lending for SMEs: Evidence From Indonesia,” *Invest. Manag. Financ. Innov.*, vol. 20, no. 2, pp. 16–25, 2023, doi: 10.21511/imfi.20(2).2023.02.

[9] A. P. Hardayu, “The Rising Popularity of Online Lending in Indonesia: The New Opportunities for Banks (Push-Pull Mooring Perspective),” *Asian J. Econ. Bus. Account.*, vol. 24, no. 6, pp. 523–539, 2024, doi: 10.9734/ajeba/2024/v24i61379.

[10] D. B. Kharisma, “Urgency of Financial Technology (Fintech) Laws in Indonesia,” *Int. J. Law Manag.*, vol. 63, no. 3, pp. 320–331, 2020, doi: 10.1108/ijlma-08-2020-0233.

[11] D. Rahmansyah, “The Role of Sharia Financial Innovation and Literacy for MSME Actors in Increasing Profitability,” *Sharia Oikonomia Law J.*, vol. 1, no. 2, pp. 126–135, 2023, doi: 10.55849/solj.v1i2.176.

[12] Y. T. Muryanto, “The Urgency of Sharia Compliance Regulations for Islamic Fintechs: A Comparative Study of Indonesia, Malaysia and the United Kingdom,” *J. Financ. Crime*, vol. 30, no. 5, pp. 1264–1278, 2022, doi: 10.1108/jfc-05-2022-0099.

[13] N. Iswari, N. Afriliana, E. Dharma, and ..., “Enhancing Aspect-based Sentiment Analysis in Visitor Review using Semantic Similarity,” *… Appl. Data Sci.*, no. Query date: 2024-10-12 10:59:38, 2024,

[14] B. Berlilana, A. M. Wahid, D. Fortuna, A. N. A. Saputra, and G. Bagaskoro, “Exploring the Impact of Discount Strategies on Consumer Ratings: An Analytical Study of Amazon Product Reviews,” *J. Appl. Data Sci.*, vol. 5, no. 1, Art. no. 1, Jan. 2024, doi: 10.47738/jads.v5i1.163.

[15] A. Paramita and T. Hariguna, “Comparison of K-Means and DBSCAN Algorithms for Customer Segmentation in E-commerce,” *J. Digit. Mark. Digit. Curr.*, no. Query

date: 2024-10-12 10:42:49, 2024,

[16] B. Hayadi and I. E. Emary, “Predicting Campaign ROI Using Decision Trees and Random Forests in Digital Marketing,” *… Digit. Mark. Digit. Curr.*, no. Query date: 2024-10-12 10:42:49, 2024,

[17] A. Hananto and D. Sugianto, “Analysis of the Relationship Between Trading Volume and Bitcoin Price Movements Using Pearson and Spearman Correlation Methods,” *J. Curr. Res. Blockchain*, no. Query date: 2024-10-12 10:44:40, 2024,

[18] T. Hariguna and A. Al-Rawahna, “Unsupervised Anomaly Detection in Digital Currency Trading: A Clustering and Density-Based Approach Using Bitcoin Data,” *J. Curr. Res. Blockchain*, no. Query date: 2024-10-12 10:44:40, 2024,

[19] C. Izumi and T. Hariguna, “In-Depth Analysis of Web3 Job Market: Insights from Blockchain and Cryptocurrency Employment Landscape,” *Int. J. Res. Metaverse*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/ijrm.v1i1.4.

[20] A. R. Yadulla, M. H. Maturi, K. Meduri, and G. S. Nadella, “Sales Trends and Price Determinants in the Virtual Property Market: Insights from Blockchain-Based Platforms,” *Int. J. Res. Metaverse*, vol. 1, no. 2, Art. no. 2, Sep. 2024, doi: 10.47738/ijrm.v1i2.9.

[21] T. Li, “Deep Learning-Powered Financial Product Recommendation System in Banks,” *J. Organ. End User Comput.*, vol. 36, no. 1, pp. 1–29, 2024, doi: 10.4018/joeuc.343257.

[22] M. B. Amnas, “FinTech and Financial Inclusion: Exploring the Mediating Role of Digital Financial Literacy and the Moderating Influence of Perceived Regulatory Support,” *J. Risk Financ. Manag.*, vol. 17, no. 3, p. 108, 2024, doi: 10.3390/jrfm17030108.

[23] K. Jangir, V. Sharma, S. Taneja, and R. Rupeika-Apoga, “The Moderating Effect of Perceived Risk on Users’ Continuance Intention for FinTech Services,” *J. Risk Financ. Manag.*, vol. 16, no. 1, p. 21, 2022, doi: 10.3390/jrfm16010021.

[24] S. Singh, “An Integrated Model Combining ECM and UTAUT to Explain Users’ Post-Adoption Behaviour Towards Mobile Payment Systems,” *Australas. J. Inf. Syst.*, 2020, doi: 10.3127/ajis.v24i0.2695.

[25] M. B. Amnas, “Understanding the Determinants of FinTech Adoption: Integrating UTAUT2 With Trust Theoretic Model,” *J. Risk Financ. Manag.*, vol. 16, no. 12, p. 505, 2023, doi: 10.3390/jrfm16120505.

[26] L. B. Benjamin, “Marketing, Communication, Banking, and Fintech: Personalization in Fintech Marketing, Enhancing Customer Communication for Financial Inclusion,” *Int. J. Manag. Entrep. Res.*, vol. 6, no. 5, pp. 1687–1701, 2024, doi: 10.51594/ijmer.v6i5.1142.

[27] C. M. Barbu, D. L. Florea, D.-C. Dabija, and M. C. R. Barbu, “Customer Experience in Fintech,” *J. Theor. Appl. Electron. Commer. Res.*, vol. 16, no. 5, pp. 1415–1433, 2021, doi: 10.3390/jtaer16050080.

[28] B. Setiawan, D. P. Nugraha, A. Irawan, R. J. Nathan, and Z. Zéman, “User Innovativeness and Fintech Adoption in Indonesia,” *J. Open Innov. Technol. Mark. Complex.*, vol. 7, no. 3, p. 188, 2021, doi: 10.3390/joitmc7030188.

[29] M. Saputra, “Consumer Behaviour and Acceptance in Fintech Adoption: A Systematic Literature Review,” *Acta Inform. Pragensia*, vol. 12, no. 2, pp. 468–489, 2023, doi: 10.18267/j.aip.222.

[30] T.-H. Cham, L. S. Cheng, L. C. Seong, A. A. Khin, and R. L. L. Bin, “Preliminary Study on Consumer Attitude Towards FinTech Products and Services in

Malaysia," *Int. J. Eng. Technol.*, vol. 7, no. 2.29, p. 166, 2018, doi: 10.14419/ijet.v7i2.29.13310.

[31] J. M. Geiger, L. M. Goodhew, and T. O. B. Odden, "Developing a Natural Language Processing Approach for Analyzing Student Ideas in Calculus-Based Introductory Physics," pp. 206–211, 2022, doi: 10.1119/perc.2022.pr.geiger.

[32] Q. D. Buchlak *et al.*, "Machine Learning Applications to Clinical Decision Support in Neurosurgery: An Artificial Intelligence Augmented Systematic Review," *Neurosurg. Rev.*, vol. 43, no. 5, pp. 1235–1253, 2019, doi: 10.1007/s10143-019-01163-8.

[33] I. Cho, M. Lee, and Y.-J. Kim, "What Are the Main Patient Safety Concerns of Healthcare Stakeholders: A Mixed-Method Study of Web-Based Text," *Int. J. Med. Inf.*, vol. 140, p. 104162, 2020, doi: 10.1016/j.ijmedinf.2020.104162.

[34] A. Fong and R. M. Ratwani, "An Evaluation of Patient Safety Event Report Categories Using Unsupervised Topic Modeling," *Methods Inf. Med.*, vol. 54, no. 04, pp. 338–345, 2015, doi: 10.3414/me15-01-0010.

[35] Y. Jiang, "Application of Game Theory Analysis in E-Commerce Platform Pricing," *Front. Bus. Econ. Manag.*, 2023, doi: 10.54097/fbem.v7i2.4361.

[36] S. Nagpal, D. Srivastava, and S. S. Mande, "What if We Perceive SARS-CoV-2 Genomes as Documents? Topic Modelling Using Latent Dirichlet Allocation to Identify Mutation Signatures and Classify SARS-CoV-2 Genomes," 2020, doi: 10.1101/2020.08.20.258772.

[37] H. Jelodar *et al.*, "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, 2018, doi: 10.1007/s11042-018-6894-4.

[38] A. Rejeb, "Exploring Blockchain Research in Supply Chain Management: A Latent Dirichlet Allocation-Driven Systematic Review," *Information*, vol. 14, no. 10, p. 557, 2023, doi: 10.3390/info14100557.

[39] C.-C. Chiu, C. Wu, T.-N. Chien, L. Kao, and J. T. Qiu, "Predicting the Mortality of ICU Patients by Topic Model With Machine-Learning Techniques," *Healthcare*, vol. 10, no. 6, p. 1087, 2022, doi: 10.3390/healthcare10061087.

[40] D. S. Maylawati, Y. J. Kumar, and F. Kasmin, "Feature-Based Approach and Sequential Pattern Mining to Enhance Quality of Indonesian Automatic Text Summarization," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 3, p. 1795, 2023, doi: 10.11591/ijeecs.v30.i3.pp1795-1804.

[41] D. A. Saputro, "Classification of Traffic Accident Information Using Machine Learning From Social Media," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 3, pp. 630–637, 2020, doi: 10.30534/ijeter/2020/04832020.

[42] R. I. Kurnia and A. S. Girsang, "Classification of User Comment Using Word2vec and Deep Learning," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 11, no. 5, pp. 1–8, 2021, doi: 10.46338/ijetae0521_01.

[43] F. M. Ridho, "Performance Analysis of Naive Bayes in Text Classification of Islamophobia Issues," *Sci. Inf. Technol. Lett.*, vol. 3, no. 1, pp. 22–30, 2022, doi: 10.31763/sitech.v3i1.1211.

[44] J. Jumadi, D. S. Maylawati, L. D. Pratiwi, and M. A. Ramdhani, "Comparison of Nazief-Adriani and Paice-Husk Algorithm for Indonesian Text Stemming Process," *Iop Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 3, p. 032044, 2021, doi: 10.1088/1757-899x/1098/3/032044.

[45] D. Mustikasari, I. Widaningrum, R. Arifin, and W. H. E. Putri, "Comparison of

Effectiveness of Stemming Algorithms in Indonesian Documents," 2021, doi: 10.2991/aer.k.210810.025.

[46] J. Luo, "Stopwords in Technical Language Processing," *Plos One*, vol. 16, no. 8, p. e0254937, 2021, doi: 10.1371/journal.pone.0254937.

[47] T. Dillan and D. H. Fudholi, "LDAViewer: An Automatic Language-Agnostic System for Discovering State-of-the-Art Topics in Research Using Topic Modeling, Bidirectional Encoder Representations From Transformers, and Entity Linking," *Ieee Access*, vol. 11, pp. 59142–59163, 2023, doi: 10.1109/access.2023.3285116.

[48] S. Khomsah and A. S. Aribowo, "Text-Preprocessing Model Youtube Comments in Indonesian," *J. Resti Rekayasa Sist. Dan Teknol. Inf.*, vol. 4, no. 4, pp. 648–654, 2020, doi: 10.29207/resti.v4i4.2035.

[49] S. Khomsah, A. F. Hidayatullah, and A. S. Aribowo, "Comparison of the Effects of Feature Selection and Tree-Based Ensemble Machine Learning for Sentiment Analysis on Indonesian YouTube Comments," pp. 161–172, 2021, doi: 10.1007/978-981-33-6926-9_15.

[50] N. Bintarawati, "Eksistensi Bahasa Indonesia Dalam Bahasa Gaul Di Universitas Veteran Bangun Nusantara," *J Indones.*, vol. 4, no. 2, p. 250, 2023, doi: 10.26858/indonesia.v4i2.49629.

[51] D. N. Djenar and M. C. Ewing, "Language Varieties and Youthful Involvement in Indonesian Fiction," *Lang. Lit. Int. J. Stylist.*, vol. 24, no. 2, pp. 108–128, 2015, doi: 10.1177/0963947015573387.

[52] E. Hendrajat, "The Impact of Slang Usage on Indonesian From a Sociolinguistic Perspective," *J. Dev. Res.*, vol. 7, no. 1, pp. 104–109, 2023, doi: 10.28926/jdr.v7i1.302.

[53] Y. Q. Yusuf, I. A. Fata, and R. Aini, "'Gotchuu!': The Use of Slang in Social Media by Generation Z," *Lingua J. Ilmu Bhs. Dan Sastra*, vol. 17, no. 2, pp. 167–174, 2023, doi: 10.18860/ling.v17i2.17756.

[54] M. A. Ramdhani, D. S. Maylawati, and T. Mantoro, "Indonesian News Classification Using Convolutional Neural Network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 2, p. 1000, 2020, doi: 10.11591/ijeecs.v19.i2.pp1000-1009.

[55] C. H. Yu, "Exploratory Data Analysis in the Context of Data Mining and Resampling.," *Int. J. Psychol. Res.*, vol. 3, no. 1, pp. 9–22, 2010, doi: 10.21500/20112084.819.

[56] J.-Y. Chang and I.-M. Kim, "Analysis and Evaluation of Current Graph-Based Text Mining Researches," pp. 100–103, 2013, doi: 10.14257/astl.2013.42.23.

[57] J. Wei and K. Zou, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks," 2019, doi: 10.18653/v1/d19-1670.

[58] S. Afzal, R. Maciejewski, Y. Jang, N. Elmqvist, and D. S. Ebert, "Spatial Text Visualization Using Automatic Typographic Maps," *Ieee Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2556–2564, 2012, doi: 10.1109/tvcg.2012.264.

[59] E. Polisciuc, A. Alves, and P. Machado, "Understanding Urban Land Use Through the Visualization of Points of Interest," 2015, doi: 10.18653/v1/w15-2810.