International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I6P132

Volume 72 Issue 6, 347-360, June 2024

Original Article

Data-Driven Insights for Mobile Banking App Improvement: A Sentiment Analysis and Topic Modelling Approach for SimobiPlus User Reviews

Edwina[1], Tuga Mauritsius[2]

[1,2]*Department of Information System Management, BINUS Graduate Program.*
Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia.

[1]*Corresponding Author : edwina@binus.ac.id*

Received: 11 March 2024 Revised: 20 May 2024 Accepted: 08 June 2024 Published: 29 June 2024

Abstract - *This study presents a comprehensive analysis of user reviews for the SimobiPlus mobile banking application in Indonesia. By leveraging state-of-the-art natural language processing techniques, including IndoBERT embeddings and machine learning classifiers (SVM, Naïve Bayes, KNN, Random Forest, Logistic Regression), we perform multi-dimensional sentiment analysis and topic modelling on a dataset of over 7,000 user reviews. Our approach classifies reviews based on sentiment (positive/negative), information type (bug report, feature request, user experience, ratings), objectives (app-related, company-related), and emotions (anger, joy, disgust, etc.). We also extract key topics and issues discussed in the reviews using Latent Dirichlet Allocation (LDA). The results demonstrate the effectiveness of SVM with hyperparameter tuning for sentiment classification (91% accuracy) and identify several recurring themes in user feedback, such as login/update errors, transaction failures, and requests for new features. Notably, we find that stemming has minimal impact on classification performance for this Indonesian language dataset. Our findings provide actionable insights for developers and managers to prioritize app improvements and enhance the overall user experience of mobile banking services. This study contributes to the growing body of research on data-driven user feedback analysis and offers practical recommendations for digital banking innovation in emerging markets.*

Keywords - *Mobile banking, Sentiment analysis, Text mining, Topic modelling.*

1. Introduction

Mobile banking represents a pivotal adaptation within the banking sector to the digital age, offering users the convenience of conducting financial transactions seamlessly through mobile devices [1]. The ease provided by mobile banking has yielded positive impacts on the trend of transaction amounts via mobile banking platforms from 2016 to 2021 in Indonesia [2]. As depicted in Figure 1, there has been a notable increase in transaction amounts, signaling mobile banking as a high-performing media delivery channel. Consequently, mobile banking presents an opportunity for banks to augment revenue streams by consistently delivering mobile banking services that effectively cater to the evolving needs of users [1].User feedback from the App Store and Google Play Store, including ratings and reviews, serves as a crucial parameter to gauge the effectiveness of mobile banking in meeting user needs [2], [3]. However, discrepancies often arise between the ratings and accompanying reviews, as illustrated in Figure 2, warranting further analysis of reviews to accurately gauge user satisfaction. These reviews, devoid of a standardized format, can be posted spontaneously by users anytime and anywhere [3], [4]. Manual analysis of hundreds of comments per month by banking institutions can result in prolonged conclusions [3], [4]. Consequently, researchers have increasingly turned to text-mining methods to expedite the review analysis process.

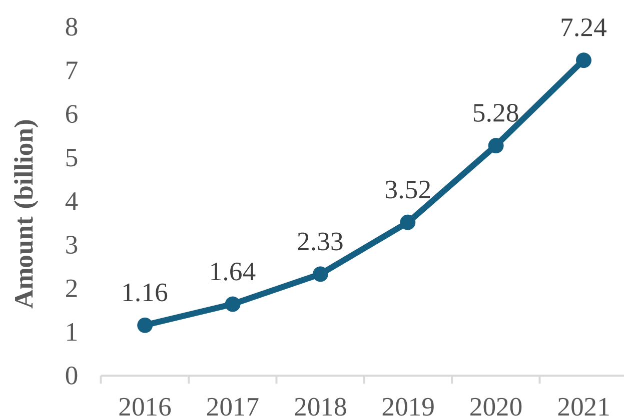

Fig. 1 Transaction Amount Using Mobile Banking [1]

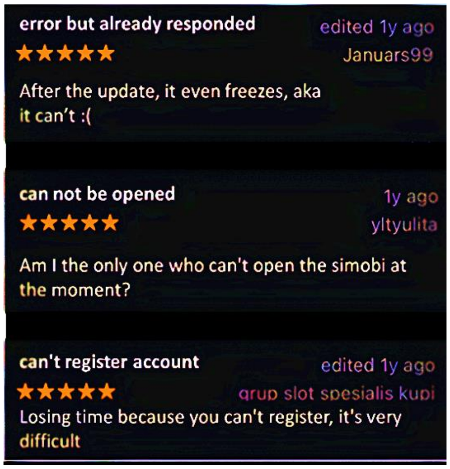

Fig. 2 Sample of rating & Review

Text mining involves extracting information from high-quality textual data and can pinpoint issues within the text pertaining to a specific topic. In the realm of sentiment analysis, text mining aids in discerning the emotional tone underlying user statements [4], [5].

Referencing the fact that mobile banking in Indonesia holds a positive appeal, investigating user reviews of mobile banking in Indonesia becomes an intriguing endeavor. This was exemplified by [6], who classified user comments on BNI Mobile Banking into positive or negative sentiments using a Support Vector Machine (SVM). With the aim of comparing algorithm performance with and without k-fold validation, they found that k-fold validation could enhance model accuracy from 78.19% to 78.45% in experiments with 60% training data and 40% testing data. [7] also delved into research to glean insights into customer satisfaction with digital banking in Indonesia (Jenius, Jago, and Blu) based on sentiment analysis from Twitter.

They employed a diverse array of algorithms, including Naïve Bayes, Logistic Regression, SVM, Decision Tree, K-Nearest Neighbors (KNN), Random Forest, Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (LGBM). SVM outperformed other algorithms with an accuracy of 74.29%. They also visualized word clouds to discern frequently discussed topics, revealing that negative sentiment often stemmed from complaints regarding poor user experience and complex policies. Furthermore, [8] conducted research to analyze user reviews of NeoBank due to its perceived low rating (3.9). Utilizing SVM, they achieved the highest accuracy of 82.33% with a scenario of 90% training data and 10% testing data. Similar to the study by [7], word cloud was also visualized in this research, highlighting that a lack of education on NeoBank app usage was the most frequently discussed negative aspect by users. From the experiment above, it is evident that SVM exhibits excellent performance in classifying mobile banking reviews in Indonesia, achieving accuracy rates exceeding 70%. To contribute to existing research, this study delves into user satisfaction with SimobiPlus, a mobile banking platform that has not been extensively studied previously. SimobiPlus presents an intriguing subject for investigation due to its substantial user base exceeding 1 million, as reported on the Google Play Store (Apr 2024). Moreover, according to [9], mobile banking usage on the SimobiPlus platform surged by 55%, with transaction values experiencing a 31% growth in 2021.

These statistics underscore SimobiPlus' potential for providing valuable insights into sentiment analysis within the Indonesian mobile banking landscape. To achieve this, we leverage various algorithms previously employed in text mining studies [6], [7], [8], including Naïve Bayes, SVM, Random Forest, Logistic Regression, and KNN. These algorithms are applied to a dataset comprising 7000 user reviews spanning the past two years (April 2021 – March 2023), sourced from both the Google Play Store and App Store. Our objective extends beyond sentiment classification (positive/negative) to encompass other dimensions, aiming for a comprehensive understanding beneficial to the company.

While previous researchers have explored various dimensions in review analysis [10], [11], [12], [13], these perspectives have not been specifically tailored to mobile banking reviews in Indonesia. Hence, our research introduces a novel approach by incorporating sentiment analysis alongside considerations for the type of information (bug report, feature request, user experience, ratings), objective (app-related & company-related), and emotions (anger, disgust, fear, joy, sadness, surprise, and neutral). Through this multifaceted approach, we aim to offer fresh insights and a deeper comprehension of user satisfaction with SimobiPlus, ultimately contributing to the enhancement of mobile banking services in Indonesia.

This study aims to identify the most suitable algorithm for classifying SimobiPlus mobile banking reviews and uncover frequently discussed topics among users. Diverging from previous research, this study employs a probability distribution-based approach to extract prevalent topics from user reviews. The algorithm utilized is Latent Dirichlet Allocation (LDA), a topic modelling machine learning method renowned for identifying patterns and relationships within text data by leveraging the concept of data distribution [14], [15], [16]. This research contributes significantly to the banking landscape in Indonesia by broadening the scope of mobile banking research objects. By providing insights into user sentiments, types of information, objectives, and emotions, as well as prevalent discussion topics, this study equips companies with valuable information for evaluating and enhancing SimobiPlus or other mobile banking applications. Ultimately, the findings of this research can drive improvements in user experience and overall service quality within the mobile banking sector.

This study is organized into several sections: in Part II, previous studies regarding the theory used in this study are presented; in Part III, the research process used in this study is explained; and in Part IV, the results of data analysis and discussion are presented. Then, in the last section, Part V will provide conclusions from the research conducted along with recommendations for future work based on this study.

2. Previous Research

2.1. Mobile Banking

Mobile banking is a form of innovation provided by banks to facilitate their customers in conducting financial transactions. According to a literature study conducted by [17], mobile banking is a service or product offered by banks using portable technology. Mobile banking that can be accessed via smartphone is a solution that can be used by customers to conduct financial transactions anytime and anywhere, as well as generate benefits in the form of time savings and shorter customer queues. Mobile banking provides benefits to the banking industry in terms of cost savings, attracting new customers and retaining old customers [17].

2.2. Text Representation

The extraction of features from unstructured data, such as textual data, is a crucial step in natural language processing. This process involves converting textual data into numerical vectors, a task that has evolved over time [18]. One traditional method for this process is Term Frequency (TF), which counts the occurrences of words in a document [7]. However, due to TF's dependency on word weights and its tendency to overlook specific words, the Term Frequency-Inverse Document Frequency (TF-IDF) method was introduced. TF-IDF combines TF with IDF to measure the importance of words in a corpus and extract significant features or keywords from text [7], [19], [20]. This method has gained popularity and is frequently utilized in sentiment analysis research [2], [3], [6], [8], [21], [22].Recently, the Bidirectional Encoder Representations from Transformers (BERT) method emerged, introduced in 2018. BERT utilizes a multi-layer transformer and operates bidirectionally on extensive datasets, incorporating a self-attention mechanism to understand the relationship between words in a sentence comprehensively [23], [24], [25]. Research conducted by [26] demonstrated that BERT outperforms TF-IDF, achieving an accuracy of 88% and highlighting its efficacy as a feature extraction method. Variations of the BERT model have been adapted for specific languages, such as IndoBERT. IndoBERT is a modification of BERT Base tailored to the Indonesian language. It is pre-trained on the Indo4B dataset, which comprises four billion pre-processed Indonesian text data sourced from various platforms, including online news, social media, Wikipedia, and more [27]. IndoBERT's pre-training process exclusively employs the Masked Language Model (MLM). This modification has been widely embraced by researchers investigating Indonesian language data.

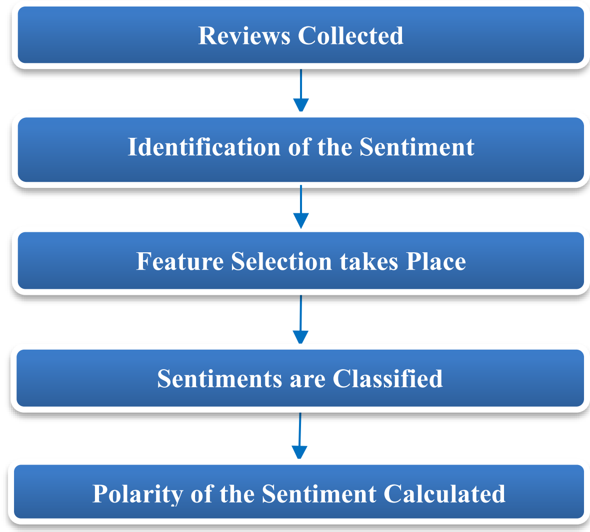

Fig. 3 Sentiment analysis process

In summary, the evolution of feature extraction methods, from TF-IDF to BERT and its variations like IndoBERT, underscores the continuous advancements in natural language processing techniques and their significant implications for research and applications in various domains.

2.3. Sentiment Analysis

Sentiment analysis is the process of scrutinizing users' sentiments or feelings regarding their experiences with an application [28], [29], [30]. This undertaking utilizes machine learning techniques, specifically the classification method derived from supervised learning. Classification stands as a cornerstone in machine learning pursuits, particularly within supervised learning, where historical data is harnessed to categorize discrete labelled data [31]. The steps involved in sentiment analysis are delineated in Figure 3.

User reviews of applications represent focal points in classification research. Typically, scholars classify these reviews into sentiment-based categories such as positive, negative, or neutral [2], [6], [7], [8], [21], [26], [32].

SVM and Naïve Bayes emerge as widely adopted algorithms renowned for their robust performance. SVM's prowess in seeking optimal hyperplanes for data segregation bolsters generalization and curbs overfitting, thereby facilitating superior data adaptation [33].

Notable examples include [7] achieving the highest accuracy in digital payment application reviews (74.29%) with SVM and [18] attaining 97.17% accuracy in mobile application reviews with SVM. Similarly, Naïve Bayes demonstrates commendable accuracy, as evidenced by [2], achieving 89.65% accuracy in mobile banking review data. Logistic Regression, KNN, and Random Forest are other frequently employed algorithms for text classification [3], [7], [21], [22], [26].

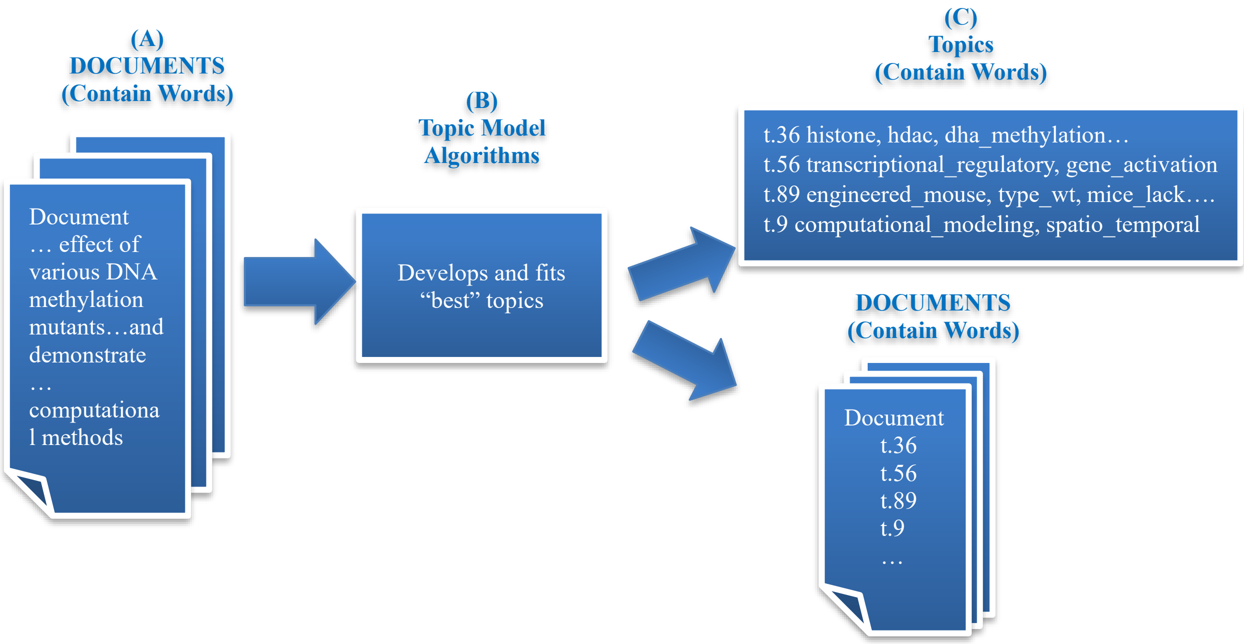

Fig. 4 Topic modelling process [34]

In further exploration of review classification, scholars can categorize reviews into additional dimensions, such as the type of information (e.g., bug report, feature request, user experience, ratings) [12], objectives (e.g., application and company) [13], and emotions (e.g., anger, disgust, fear, joy, sadness, surprise, neutral) [10].

Beyond merely gauging user sentiment toward the application, this classification facilitates swift identification of the reviewer's intent, target audience, and emotional response to using the application, thereby furnishing invaluable insights for product enhancement and user satisfaction.

2.4. Topic Modelling

Topic modelling emerges as a potent tool for scrutinizing extensive volumes of textual data, including content sourced from social media platforms [35]. It functions on the premise that each document comprises a blend of topics, with each topic delineated by a probability distribution of words [36]. Latent Dirichlet Allocation (LDA) stands at the forefront of topic modelling techniques in machine learning, aimed at uncovering intricate patterns and correlations within textual documents [14], [15], [16].

This method relies on probability distributions, positing that a distribution of topics can elucidate each document while each topic can be explicated by a distribution of words [32].

Within the realm of topic modelling, the algorithm operates by calculating the number of topics, each labelled based on when the highest coherence value is attained. A higher coherence value signifies the optimal selection of topics. Interpretation of each topic is deduced based on the weight/probability of each word within the respective topic. A higher probability value indicates a stronger association of the word with the topic.

Researchers commonly harness review data to embark on topic modelling endeavours, with the objective of uncovering pivotal themes and insights. For instance, [32] employed LDA to delve into topics within user reviews of mobile banking applications, thereby pinpointing areas necessitating enhancement. They highlighted elements such as convenience, simplicity, and usability as commendable features to maintain while shedding light on issues like OTP code transmission failures, connectivity glitches, and challenges encountered during the login and registration process, warranting attention.

Beyond expediting review classification, topic extraction from user reviews assumes paramount importance for conducting comprehensive analyses. This approach empowers companies to delve deeper into the strengths and weaknesses of their applications, thereby facilitating continuous assessment and refinement aimed at augmenting user satisfaction [32]. The topic modelling process is visually depicted in Figure 4.

3. Methodology

The Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which is an open standard process model that describes the general approach used by data mining experts [37], [38], was used in this study. The steps taken can be seen in Figure 5.

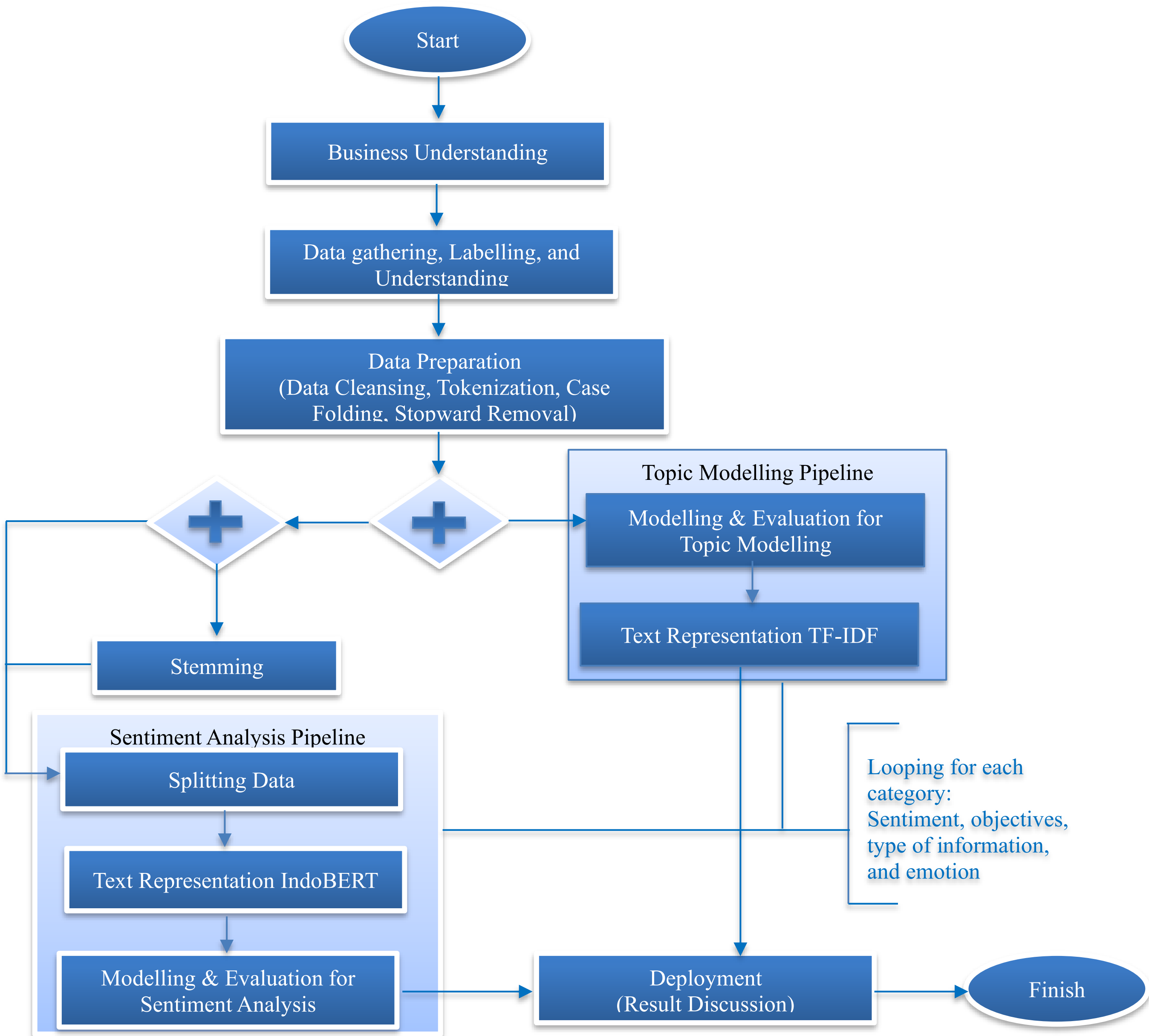

Fig. 5 Research flow

3.1 Business Understanding

The Business Understanding process is carried out to determine the problem and final goal of the research from the business side. Seeing the fact that mobile banking is a promising business in the banking industry encourages banks to continue providing mobile banking services that can meet the needs of its users.

Information mining from SimobiPlus user review data using a machine learning approach was carried out in this research so that banks can process review data on a large scale quickly and focus on creating business process innovations that are relevant to the needs of their users in accordance with the input received.

3.2. Data Understanding

SimobiPlus user review data spanning from April 2021 to March 2023 was scraped from both the Google Play Store and App Store using Python version 3.9. Utilizing the google_play_store and app_store_scraper libraries, a total of 7,017 reviews were collected, comprising 232 from the App Store and 6,785 from the Google Play Store.

Each review record was given 4 label categories, a process facilitated by three individuals who were not language experts. This approach was adopted due to the informal nature of the reviews, written in everyday language easily interpretable by the public accustomed to casual language use. The 4 categories and distribution of labels can be seen in Figure 6.

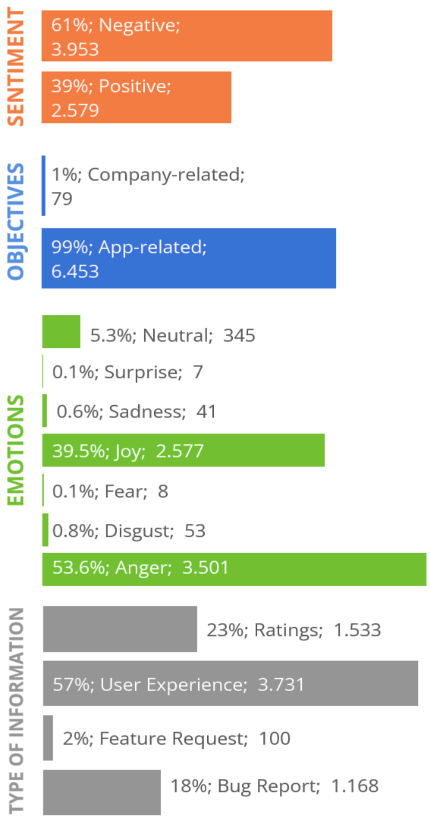

Fig. 6 Label distribution

When carrying out the data labeling process, 147 reviews were found whose meaning could not be interpreted, such as "Kkfx xx Szzza" and "Jx". 294 full reviews in English are also marked because this research focuses on processing Indonesian language review data. Reviews that could not be interpreted or were in full English were not removed from the dataset because it is inevitable that there will be similar reviews in the future. These two review categories are given fuzzy labels so that if there is a similar review later, the model can still have optimal performance because it can handle reviews other than reviews in Indonesian.

3.3. Data Preparation

Referring to research conducted by [6], 6 stages were carried out in text data preparation. The 6 stages consist of:

1. Data cleaning: clean data from punctuation marks, numbers, and spaces, and translate emojis into words.
2. Tokenization: breaking down sentences or groups of sentences into words, symbols, or other elements that have meaning.
3. Case folding: converts all characters to lowercase.
4. Spelling Normalization: changing the spelling of a word to a more common or correct standard form.
5. Stopword Removal (filtering): delete words that appear frequently and do not have important meaning.
6. Stemming: the process of changing words back into basic words.

Table 1. Example of data preparation

Original Review	Veryyyyyy good apk like it 😅
Data Cleaning	Veryyyyyy good apk like it face smile and sweat
Tokenization	['Veryyyyyy', 'good', 'apk', 'like', 'it', 'face', 'smile', 'and', 'sweat']
Case Folding	['veryyyyyy', 'good', 'apk', 'like', 'it', 'face', 'smile', 'and', 'sweat']
Spelling Normalization	['veryyyyyy', 'good', 'application, 'like', 'it', 'face', 'smile', 'and', 'sweat']
Stopword Removal	['veryyyyyy', 'good', 'application, 'like', 'face', 'smile', 'sweat']
Stemming	['very, 'good', 'application, 'like', 'face', 'smile', 'sweat']

These 6 stages can be seen in Table 1.

Furthermore, based on previous research [39], where affixes can give different meanings to a word in Indonesian, in this research, two types of experiments were carried out, which is:

(a) Non-Stemming Dataset Experience: using datasets without stemming (only carrying out 5 stages of data preparation).
(b) Stemming Dataset Experience: using datasets that go through the stemming step.

Then, the process was broken down into 2 parts, such as review classification into multi-dimensional sentiment analysis and topic modelling.

3.3.1. Multi-Dimensional Sentiment Analysis

In this section, the process of data separation and feature extraction is carried out. The two datasets (Non-Stemming Dataset and Stemming Dataset) were then divided in a ratio of 80:20 for training data and test data, where the percentage refers to the Pareto Principle [40].

In this data-splitting process, the stratify parameter is used, which helps ensure that the class proportions in the initial data are maintained in both subsets of the resulting data (in the training data and testing data).

This parameter is crucial for handling class imbalance because it helps prevent the training or testing data from being skewed towards one class, which could lead to biased model performance.

By maintaining a proportional representation of each class in both sets, the model learns more effectively from diverse examples and can make more accurate predictions for minority classes [41]. The split dataset is then extracted into a numerical vector using the text representation of theIndoBERT pre-train model.

Table 2. Hyperparameter Configuration

ALGORITHM	PARAM
SVM (kernel: RBF)	param_grid = {'estimator__C': [0.1, 1, 10], 'estimator__gamma': [0.001, 0.01, 0.1, 1]}
SVM (kernel: Poly)	param_grid = {'estimator__C': [0.1, 1, 10], 'estimator__degree': [2, 3, 4]}
Random Forest	param_grid = {'estimator__bootstrap': [True], 'estimator__max_features': [2, 3], 'estimator__max_depth': [20, 30], 'estimator__min_samples_leaf': [2, 4], 'estimator__min_samples_split': [5, 10], 'estimator__n_estimators': [100, 200]}
Naïve Bayes	param_grid = {'var_smoothing': [1e-9, 1e-8, 1e-7, 1e-6, 1e-5, 1e-4]}
Logistic Regression	param_grid = {'estimator__C': [0.001, 0.1, 1], 'estimator__penalty': ['l1', 'l2'], 'estimator__solver' : ['liblinear', 'lbfgs', 'newton-cg']}
KNN	param_grid = {'estimator__n_neighbors': [3, 5, 7], 'estimator__weights': ['uniform', 'distance']}

3.3.2. Topic Modelling

In this section, the feature extraction process is carried out. The dataset used in the topic modelling experiment is the Non-Stemming Dataset. The feature extraction process in this experiment was carried out using TF-IDF because the topic modelling algorithm that will be used (LDA) does not yet support BERT text representation.

3.4. Modelling

Continuing the results of text representation in the previous sub-chapter, the modelling process is also divided into 2 parts, such as multi-dimensional sentiment analysis and topic modelling.

3.4.1. Multi-dimensional Sentiment Analysis

The training phase on the dataset employs various machine learning algorithms, including Naïve Bayes, SVM (with RBF and Polynomial kernels), Logistic Regression, Random Forest, and KNN, to construct a review prediction model based on initial labels. Data balancing is achieved using the class_weight = ‘balanced’ parameter in each algorithm, which automatically assigns greater weight to minority classes, thereby enhancing the model’s ability to handle class imbalances without the introduction of synthetic data, which may add complexity.

However, this parameter is not utilized by Naïve Bayes due to its inherent limitations. In this study, a total of 96 sentiment analysis models were developed, encompassing 4 labels associated with SimobiPlus mobile banking user reviews. These models represent a fusion of 6 algorithms with diverse treatments, including some employing stemming, while others do not, some undergoing hyperparameter tuning, and others not. The hyperparameter configurations utilized are detailed in Table 2.

3.4.2. Topic Modeling Experiment

A subsequent experiment was conducted to harness the dataset for crafting a topic model through Latent Dirichlet Allocation (LDA). This model is instrumental in discerning prevalent themes within the review dataset under examination.

Initially, 21 topics were initiated to unravel recurring subjects deliberated by users. Subsequently, during the Evaluation phase, pertinent topics will be meticulously sieved for further analysis.

3.5. Evaluation

In the comprehensive sentiment analysis framework, various evaluation metrics were employed to assess the efficacy of each algorithm utilized. The performance of each algorithm is detailed in Table 3.

Additionally, a comprehensive elucidation of each evaluation metric utilized is provided in the following exposition [42], [43], [44].

1. Accuracy: the value of all correct predictions for all predictions produced.
2. F1-Score: harmonic mean of precision and recall (this metric is useful when balancing precision and recall).
3. Precision: TP prediction value for all positive predictions.
4. Recall: TP predicted the value of all actual positives.

Additionally, several terms are employed to assess the performance of an algorithm [44], [45]:

1. True positive (TP): the prediction indicates a positive outcome and aligns with the actual positive classification.
2. False positive (FP): The prediction indicates a negative outcome, yet the actual classification is positive.
3. False negative (FN): The prediction indicates a positive outcome, but the actual classification is negative.

Table 3. Evaluation results of review classifications

	Method		Non Hyperparameter Tuning				Hyperparameter Tuning			
			F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall
Sentiment	SVM (kernel: RBF)	Non-Stemming	0.83	0.83	0.84	0.83	0.91	0.91	0.91	0.91
	SVM (kernel: Poly)		0.84	0.84	0.85	0.84	0.89	0.89	0.90	0.89
	Logistic Regression		0.90	0.90	0.91	0.90	0.90	0.90	0.91	0.90
	Random Forest		0.88	0.88	0.88	0.88	0.87	0.87	0.87	0.87
	KNN		0.83	0.83	0.83	0.83	0.87	0.87	0.87	0.87
	Naive Bayes		0.67	0.66	0.71	0.66	0.67	0.66	0.71	0.66
	SVM (kernel: RBF)	Stemming	0.83	0.83	0.84	0.83	0.91	0.91	0.91	0.91
	SVM (kernel: Poly)		0.84	0.84	0.85	0.84	0.89	0.89	0.90	0.89
	Logistic Regression		0.90	0.90	0.91	0.90	0.90	0.90	0.91	0.90
	Random Forest		0.88	0.88	0.88	0.88	0.87	0.87	0.87	0.87
	KNN		0.84	0.84	0.85	0.84	0.87	0.87	0.87	0.87
	Naive Bayes		0.67	0.66	0.71	0.66	0.67	0.66	0.71	0.66
Objectives	SVM (kernel: RBF)	Non-Stemming	0.90	0.87	0.94	0.87	0.95	0.96	0.95	0.96
	SVM (kernel: Poly)		0.89	0.86	0.94	0.86	0.92	0.91	0.95	0.91
	Logistic Regression		0.93	0.92	0.95	0.92	0.93	0.92	0.95	0.92
	Random Forest		0.95	0.96	0.95	0.96	0.96	0.96	0.95	0.96
	KNN		0.92	0.91	0.94	0.91	0.95	0.96	0.95	0.96
	Naive Bayes		0.35	0.27	0.91	0.27	0.35	0.27	0.91	0.27
	SVM (kernel: RBF)	Stemming	0.90	0.87	0.94	0.87	0.95	0.96	0.95	0.96
	SVM (kernel: Poly)		0.89	0.86	0.94	0.86	0.92	0.91	0.95	0.91
	Logistic Regression		0.93	0.92	0.95	0.92	0.93	0.92	0.95	0.92
	Random Forest		0.95	0.96	0.95	0.96	0.96	0.96	0.95	0.96
	KNN		0.92	0.91	0.94	0.91	0.94	0.93	0.94	0.93
	Naive Bayes		0.35	0.27	0.91	0.27	0.35	0.27	0.91	0.27
Type of Information	SVM (kernel: RBF)	Non-Stemming	0.72	0.70	0.74	0.70	0.75	0.76	0.76	0.76
	SVM (kernel: Poly)		0.72	0.71	0.75	0.71	0.74	0.73	0.76	0.73
	Logistic Regression		0.74	0.73	0.77	0.73	0.74	0.73	0.77	0.73
	Random Forest		0.69	0.74	0.71	0.74	0.71	0.75	0.73	0.75
	KNN		0.64	0.62	0.70	0.62	0.71	0.72	0.71	0.72
	Naive Bayes		0.36	0.35	0.56	0.35	0.36	0.35	0.56	0.35
	SVM (kernel: RBF)	Stemming	0.72	0.70	0.74	0.70	0.75	0.76	0.76	0.76
	SVM (kernel: Poly)		0.72	0.71	0.75	0.71	0.74	0.73	0.76	0.73
	Logistic Regression		0.74	0.73	0.77	0.73	0.74	0.73	0.77	0.73
	Random Forest		0.69	0.74	0.71	0.74	0.71	0.75	0.73	0.75
	KNN		0.65	0.63	0.71	0.63	0.67	0.66	0.69	0.66
	Naive Bayes		0.36	0.35	0.56	0.35	0.36	0.35	0.57	0.35
Emotion	SVM (kernel: RBF)	Non-Stemming	0.71	0.66	0.80	0.66	0.83	0.84	0.84	0.84
	SVM (kernel: Poly)		0.72	0.67	0.80	0.67	0.80	0.78	0.78	0.78
	Logistic Regression		0.81	0.80	0.83	0.80	0.81	0.80	0.80	0.80
	Random Forest		0.78	0.80	0.77	0.80	0.77	0.80	0.80	0.80
	KNN		0.71	0.67	0.77	0.67	0.77	0.79	0.79	0.79
	Naive Bayes		0.22	0.23	0.78	0.23	0.22	0.23	0.23	0.23
	SVM (kernel: RBF)	Stemming	0.71	0.66	0.80	0.66	0.83	0.84	0.83	0.84
	SVM (kernel: Poly)		0.72	0.67	0.80	0.67	0.80	0.78	0.82	0.78
	Logistic Regression		0.81	0.80	0.83	0.80	0.81	0.80	0.83	0.80
	Random Forest		0.78	0.80	0.77	0.80	0.77	0.80	0.76	0.80
	KNN		0.70	0.66	0.77	0.66	0.74	0.72	0.76	0.72
	Naive Bayes		0.22	0.23	0.78	0.23	0.22	0.22	0.78	0.22

In the hyperparameter tuning experiments, the performance metrics showcased in Table 3 are derived from employing hyperparameters that yield optimal performance. The quest for optimal parameters utilizes Python's Grid Search library. The resulting optimal parameter combinations for each category are outlined in Table 4. In the context of topic modelling experiments, the coherence value serves as a pivotal metric for evaluating the resultant topic models. Determining the optimal number of topics is guided by the coherence score, which gauges the semantic cohesion of words within each topic [46]. A higher coherence score signifies a more robustly constructed model [46]. Illustrated in Figure 7, the highest coherence value within the negative class is 0.59, achieved with 16 topics. Table 5 presents three out of the 16 topics observed, along with their distribution and interpretation within the negative class. Further insights into the assessment outcomes of review classification and topic modelling are expounded upon in Section 4. To identify actionable feedback for enhancing the application, the top three topics were extracted from five distinct classes: negative, application, bug reports, anger, and positive. The topics, along with their interpretations, can be found in Table 5 (negative), Table 6 (application), Table 7 (bug reports), Table 8 (anger), and Table 9 (positive), respectively. These insights provide valuable direction for refining the application based on user feedback.

Table 4. Optimal hyperparameter configuration

Category	Best Algorithm	Hyperparameter
Sentiment	SVM (kernel: RBF)	{'estimator__C': 10, 'estimator__gamma': 0.01}
Objectives	Random Forest	{'estimator__bootstrap': True, 'estimator__max_depth': 20, 'estimator__max_features': 3, 'estimator__min_samples_leaf': 2, 'estimator__min_samples_split': 10, 'estimator__n_estimators': 200}
Type of Information	SVM (kernel: RBF)	{'estimator__C': 10, 'estimator__gamma': 0.1}
Emotions	SVM (kernel: RBF)	{'estimator__C': 10, 'estimator__gamma': 0.01}

Table 5. Examples of topics and the interpretation for negative class

1	0.055*"sms" + 0.041*"token" + 0.026*"send" + 0.019*"convenient" + 0.012*"feeling" + 0.011*"enter" + 0.010*"server" + 0.009*"free" + 0.009* "cost" + 0.009*"troublesome"	**Interpretation:** Customers feel inconvenienced because tokens are sent via SMS and charges are charged to the customer.
2	0.057*"update" + 0.041*"can" + 0.036*"no" + 0.033*"application" + 0.030*"why" + 0.024*"open" + 0.021*"difficult" + 0.019*"error" + 0.019* "simobi" + 0.018*"login"	**Interpretation:** Cannot update an error when logging in.
3	0.033*"password" + 0.031*"noise" + 0.031*"pin" + 0.030*"true" + 0.028*"blocked" + 0.020*"bad" + 0.019*"false" + 0.019*"blocked" + 0.018* "reset" + 0.017*"enter"	**Interpretation:** The account is blocked/you have to reset the password to enter the application because the pin/password is detected incorrectly, but the customer feels it is correct.

Table 6. Examples of topics and the interpretation for application class

1	4, 0.055*"fast" + 0.035*"stars" + 0.035*"like" + 0.031*"makes it easy" + 0.031*"love" + 0.029*"transaction" + 0.028*"makes it easy" + 0.024*"smiles" + 0.024*"free" + 0.024*"add"	**Interpretation:** Feel satisfied because the application can make transactions easier.
2	3, 0.036*"error" + 0.033*"login" + 0.031*"bad" + 0.030*"login" + 0.028*"no" + 0.027*"transfer" + 0.025*"bank" + 0.025*"difficult" + 0.023*"credit" + 0.022*"balance"	**Interpretation:** Error when logging in and difficulty in making a transfer.
3	11, 0.094*"update" + 0.073*"can" + 0.064*"no" + 0.062*"application" + 0.044*"why" + 0.043*"easy" + 0.043*"transaction" + 0.040*"simobi" + 0.039*"open" + 0.029*"complicated"	**Interpretation:** Difficult to open the application and make transactions after the update.

Table 7. Examples of topics and the interpretation for bug reports class

1	0.009*"claim" + 0.005*"morning" + 0.005*"gift" + 0.004*"reward" + 0.004* "difficult" + 0.004*"urgent" + 0.004*"hang" + 0.003*"shopeepay" + 0.003* "maintenance" + 0.003*"shopee"	**Interpretation:** Difficult to claim rewards.
2	0.025*"update" + 0.022*"can" + 0.020*"why" + 0.019*"open" + 0.017*"application" + 0.017*"no" + 0.017*"error" + 0.013*"balance" + 0.013* "login" + 0.012*"failed"	**Interpretation:** Can't update and can't open the application/login.
3	0.006*"face" + 0.006*"send" + 0.006*"bill" + 0.006*"open" + 0.006*"credit" + 0.006*"error" + 0.005*"problem" + 0.005*"pay" + 0.005* "like" + 0.005*"top"	**Interpretation:** An error occurred when sending the bill payment.

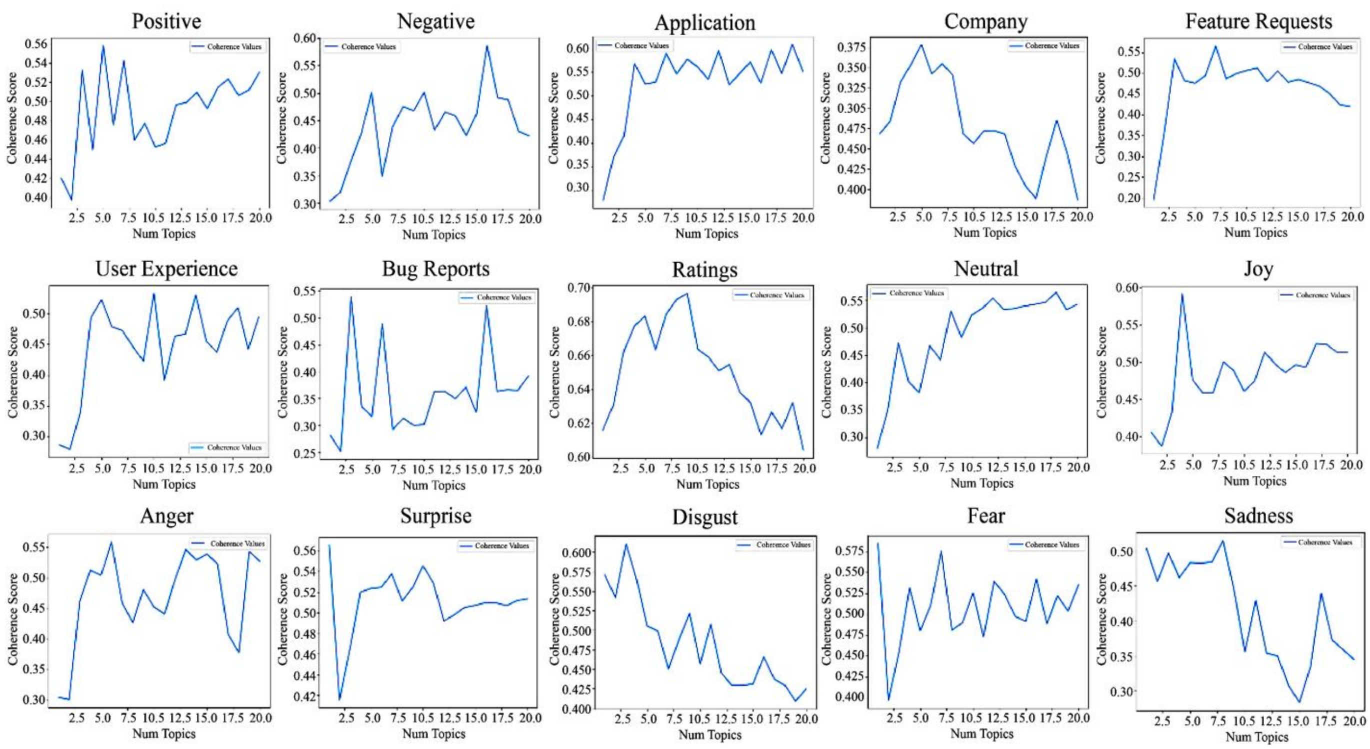

Fig. 7 Coherence values

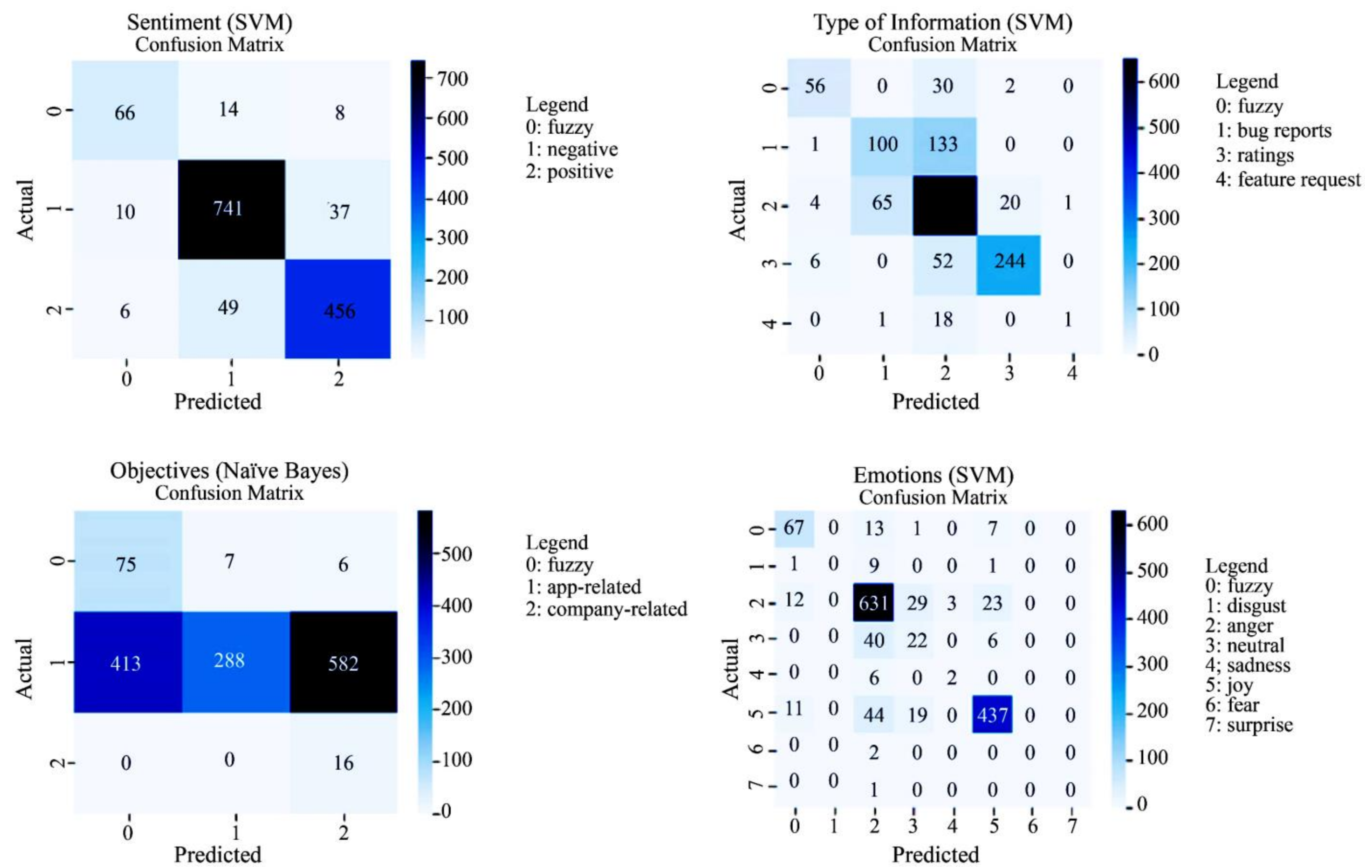

Fig. 8 Confusion matrix

Table 8. Examples of topics and the interpretation for anger class

1	0.017*"frequency" + 0.017*"update" + 0.016*"claim" + 0.007*"change" + 0.006*"check" + 0.006*"display" + 0.006*"always" + 0.006*"play" + 0.006* "selfie" + 0.005*"hopefully"	**Interpretation:** Too frequent application updates.
2	0.057*"update" + 0.038*"can" + 0.029*"no" + 0.029*"why" + 0.026* "application" + 0.022*"open" + 0.019*"error" + 0.017*"simobi" + 0.016* "balance" + 0.016*"bad"	**Interpretation:** Cannot open the application and check the balance after updating.
3	0.014*"true" + 0.013*"register" + 0.011*"account" + 0.011*"pin" + 0.010* "failed" + 0.010*"download" + 0.010*"login" + 0.010*"complicated" + 0.010* "app" + 0.009*"bank"	**Interpretation:** The pin entered was correct and the same as when registered but failed to enter.

Table 9. Examples of topics and the interpretation for the positive class

1	0.039*"easy" + 0.033*"satisfying" + 0.031*"application" + 0.028*"convenient" + 0.028*"fair" + 0.027*"top" + 0.025*"safe" + 0.020*"transfer" + 0.019* "free" + 0.013*"update"	**Interpretation:** Feel easy, satisfied, and comfortable in using the application. Happy with the Free Transfer Fee service provided by the bank.

3.6. Deployment

Deployment in this research is a process of analyzing the results and evaluating the review classification and topic selection. A description of the analysis can be seen in Section 4. Interpretation of each result obtained and important matters relating to the company's business side can be seen in this section. Recommendations will also be provided that are easy for companies to understand so that they can be used as analytical material that can support companies in making decisions regarding the development of the SimobiPlus mobile banking application.

4. Result and Discussion

This section presents a thorough discussion of the experimental findings. Upon examining the evaluation metrics outlined in Table 3, disparities from prior research [39] emerge, indicating that employing stemming in processing SimobiPlus mobile banking user reviews yield no significant impact on model performance. Remarkably, across all categories, the accuracy values of algorithms with optimal performance remain consistent between experiments with and without stemming. Moreover, the observed higher F1-Score, Accuracy, Precision, and Recall metrics in experiments incorporating hyperparameter tuning underscore its efficacy in enhancing model performance.

When examining the model performance depicted in Table 3 across sentiment analysis experiments, SVM (kernel: RBF) emerges as the top performer, achieving the highest F1-Score and accuracy rates of 91% for sentiment labels, 76% for types of information, and 84% for emotions. Remarkably, this echoes findings from prior research [6], [7], [8], emphasizing SVM's consistent efficacy in attaining commendable accuracy levels. Notably, for label classification tasks, Random Forest demonstrates superior performance, boasting an impressive F1-Score and accuracy of 96%. Conversely, Naïve Bayes falls short in accuracy due to its lack of parameters for data balancing, such as class_weight = 'balanced', setting it apart from other algorithms. We also visualize the confusion matrix of the algorithms deemed noteworthy for discussion within each category in Figure 8. Several key highlights from the visualizations include:

- The True Positive (TP) value for sentiment categories (1,263) surpasses both False Positives (FP) and False Negatives (FN) (124). This indicates the model's success in minimizing prediction errors by keeping FP and FN rates at a low level.
- In the objectives category, Naïve Bayes lacks parameters for equalizing weights in imbalanced testing data. Consequently, it yields significantly higher precision compared to other evaluation metrics (resulting in too many app-related classes predicted as company-related classes).
- TP values (897) exceed both FN and FP in the type of information category (user experience and rating classes). This model excels in predicting these two classes.
- TP values (1,068) also surpass FN and FP in the emotions category (joy and angry classes), indicating the model's proficiency in predicting these two classes.

Further evaluation of the emotional classes implemented in this study is also imperative. Upon examination of Figures 6 and 8, it becomes evident that the angry and happy classes overwhelmingly dominate the spectrum of emotions expressed in SimobiPlus user reviews. This imbalance results in the model's overfitting to these two classes, consequently diminishing its sensitivity in predicting other classes.In addition to addressing one of the primary objectives of this research, which involves identifying areas frequently reviewed for company evaluation and improvement, the analysis will focus on topics pertinent to user complaints and application development across Positive (Table 9), Negative (Table 5), Application (Table 6), Bug Reports (Table 7), and

Anger (Table 8) categories. Among the myriad topics discussed by users, updates emerge as the foremost concern, often leading to errors upon app launch or while checking balances post-update. Users express frustration over frequent memory and data-intensive update prompts. Additionally, account blocks due to PIN errors despite accurate entries and errors during point redemption are prevalent issues highlighted in bug reports and anger-related discussions. Moving beyond application improvement requests, a recurring theme in corporate discourse revolves around perceived delays in the bank's responsiveness to customer issues. Furthermore, grievances related to OTP authentication in the disgust category underscore user concerns. Users reliant on Wi-Fi/data packages lacking phone credit encounter hurdles accessing specific app features, thereby disrupting their activities. Moreover, Shopeepay and Dana, two prominent digital payment platforms, emerge as focal points in the feature request category, indicating user demand for their inclusion within the SimobiPlus application. In the positive domain, free transfer fees emerge as the most discussed aspect, evoking user satisfaction. The fee-free transfer feature via Bi-Fast offered by the SimobiPlus application serves as a notable advantage, motivating users to conduct transactions through SimobiPlus mobile banking.

From the elucidation of frequently discussed topics, several practical implications emerge as valuable lessons for banks or developers:

1. Enhancing the current system to prevent errors upon application launch and balance checking post-update.
2. Being more responsive in addressing customer issues.
3. Considering alternative authentication methods to replace OTP usage within the application flow.
4. Sustaining the Free Transfer Fee feature.

These recommendations offer actionable insights for improving user experience and optimizing the functionality of the application.

5. Conclusion

In conclusion, this study demonstrates the value of multi-dimensional sentiment analysis and topic modelling for understanding user perspectives on mobile banking applications. By leveraging advanced NLP techniques and a large-scale dataset of user reviews, we identified key drivers of user satisfaction and dissatisfaction with the SimobiPlus app in Indonesia. Our results highlight the importance of stable app performance, seamless login and update processes, and continuous feature innovation for meeting evolving customer needs. Methodologically, our comparison of various machine learning algorithms and text representations (e.g., stemming vs. non-stemming, hyperparameter tuning vs non-hyperparameter tuning) provides practical insights for researchers and practitioners working on similar text classification tasks in Indonesian or other low-resource languages. The strong performance of SVM with hyperparameter tuning and the minimal impact of stemming offer valuable benchmarks for future studies. However, our study also has several limitations that should be addressed in future work. The reliance on a single app and time period may limit the generalizability of the findings to other mobile banking contexts. Additionally, the manual labeling of review categories (e.g., information type, emotions) introduces some subjectivity and potential inconsistencies. Future research could explore more sophisticated methods for automated review categorization and sentiment scoring.

Acknowledgements

We would like to thank BINUS University who has supported our work, particularly the BINUS Graduate Program, Master of Management Information System.

References

[1] Payment System Statistics and Financial Market Infrastructure (SPIP), Bank Indonesia. [Online]. Available: https://www.bi.go.id/id/statistik/ekonomi-keuangan/ssp/transaksi-delivery-channel.aspx.

[2] Salah Al-Hagree, and Ghaleb Al-Gaphari, "Arabic Sentiment Analysis Based Machine Learning for Measuring User Satisfaction with Banking Services' Mobile Applications: Comparative Study," *2022 2nd International Conference on Emerging Smart Technologies and Applications*, Ibb, Yemen, pp. 1-4, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yudo Ekanata, and Indra Budi, "Mobile Application Review Classification for the Indonesian Language Using Machine Learning Approach," *2018 4th International Conference on Computer and Technology Applications*, Istanbul, Turkey, pp. 117-121, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Said A. Salloum et al., "A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives," *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 1, pp. 127-133, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Rachmawan Adi Laksono et al., "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor Using Naïve Bayes," *2019 12th International Conference on Information & Communication Technology and System*, Surabaya, Indonesia, pp. 49-54, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yuni Handayani, Alvin Rinaldy Hakim, and Muljono, "Sentiment Analysis of Bank BNI User Comments Using the Support Vector Machine Method," *2020 International Seminar on Application for Technology of Information and Communication*, Semarang, Indonesia, pp. 202-207, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Bramanthyo Andrian et al., "Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 3, pp. 466-473, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kusnawi Kusnawi, Majid Rahardi, and Van Daarten Pandiangan, "Sentiment Analysis of Neobank Digital Banking Using Support Vector Machine Algorithm in Indonesia," *International Journal on Informatics Visualization*, vol. 7, no. 2, pp. 377-383, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] M. Richard, Pengguna Mobile Banking Bank Sinarmas (BSIM) Melonjak 55 Persen Tahun Ini, Finansial, 2021. [Online]. Available: https://finansial.bisnis.com/read/20210823/90/1432719/pengguna-mobile-banking-bank-sinarmas-bsim-melonjak-55-persen-tahun-ini

[10] Paul Ekman, and Daniel Cordaro, "What is Meant by Calling Emotions Basic," *Emotion Review*, vol. 3, no. 4, pp. 364-370, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[11] Hao Fei et al., "Latent Emotion Memory for Multi-Label Emotion Classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 5, pp. 7692-7699, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Walid Maalej, and Hadeer Nabil, "Bug Report, Feature Request, or Simply Praise? on Automatically Classifying App Reviews," *2015 IEEE 23rd International Requirements Engineering Conference*, Ottawa, ON, Canada, pp. 116-125, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Rein Rachman Putra, Monika Evelin Johan, and Emil Robert Kaburuan, "A Naïve Bayes Sentiment Analysis for Fintech Mobile Application User Review in Indonesia," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 5, pp. 1856-1860, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Moh. Nasrul Aziz et al., "Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction," *2018 5th International Conference on Information Technology, Computer, and Electrical Engineering*, Semarang, Indonesia, pp. 125-130, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Gede Rizky Gustisa Wisnu et al., "Sentiment Analysis and Topic Modelling of 2018 Central Java Gubernatorial Election Using Twitter Data," *2020 International Workshop on Big Data and Information Security*, Depok, Indonesia, pp. 35-40, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Rajesh Prabhakar Kaila, and A.V. Krishna Prasad, "Informational Flow on Twitter – Corona Virus Outbreak – Topic Modelling Approach," *International Journal of Advanced Research in Engineering and Technology*, vol. 11, no. 3, pp. 128-134, 2020. [Google Scholar] [Publisher Link]

[17] Carlos Tam, and Tiago Oliveira, "Does Culture Influence M-Banking Use and Individual Performance?," *Information and Management*, vol. 56, no. 3, pp. 356-363, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Nurazzah Abd Rahman, Seri Dahlia Idrus, and Noor Latiffah Adam, "Classification of Customer Feedbacks Using Sentiment Analysis towards Mobile Banking Applications," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 4, pp. 1579-1587, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Eman Alsagour, Lubna Alhenki, and Mohammed Al-Dhelaan, "Different Word Representation for Text Classification: A Comparative Study," *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications*, Abu Dhabi, United Arab Emirates, pp. 1-2, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Chenrui Lu et al., "A Document Analysis of Peak Carbon Emissions and Carbon Neutrality Policies Based on a PMC Index Model in China," *International Journal of Environmental Research and Public Health*, vol. 19, no. 15, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Hilman Wisnu, Muhammad Afif, and Yova Ruldevyani, "Sentiment Analysis on Customer Satisfaction of Digital Payment in Indonesia: A Comparative Study Using KNN and Naïve Bayes," *Journal of Physics: Conference Series*, vol. 1444, no. 1, pp. 1-10, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Sakshi Ranjan, and Subhankar Mishra, "Comparative Sentiment Analysis of App Reviews," *2020 11th International Conference on Computing, Communication and Networking Technologies*, Kharagpur, India, pp. 1-7, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Jacob Devlin et al., "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, Minneapolis, Minnesota, pp. 4171-4186, 2019. [Google Scholar] [Publisher Link]

[24] Muchammad Naseer, Muhamad Asvial, and Riri Fitri Sari, "An Empirical Comparison of BERT, RoBERTa, and Electra for Fact Verification," *2021 International Conference on Artificial Intelligence in Information and Communication*, Jeju Island, Korea (South), pp. 241-246, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Anwar Hussen Wadud et al., "Deep-BERT: Transfer Learning for Classifying Multilingual Offensive Texts on Social Media," *Computer Systems Science and Engineering*, vol. 44, no. 2, pp. 1775-1791, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Saeid Pourroostaei Ardakani et al., "A Data-Driven Affective Text Classification Analysis," *2021 20th IEEE International Conference on Machine Learning and Applications*, Pasadena, CA, USA, pp. 199-204, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] Sani Muhamad Isa, Gary Nico, and Mikhael Permana, "Indobert for Indonesian Fake News Detection," *ICIC Express Letters*, vol. 16, no. 3, pp. 289-297, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Sebastian Kopera et al., "Interdisciplinarity in Tech Startups Development - Case Study of 'Unistartapp' Project," *Foundations of Management*, vol. 10, no. 1, pp. 23-32, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[29] Jose Ramon Saura, Ana Reyes-Menendez, and Cesar Alvarez-Alonso, "Do Online Comments Affect Environmental Management? Identifying Factors Related to Environmental Management and Sustainability of Hotels," *Sustainability*, vol. 10, no. 9, pp. 1-20, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[30] Jose Ramon Saura, Pedro Palos-Sanchez, and Antonio Grilo, "Detecting Indicators for Startup Business Success: Sentiment Analysis Using Text Data Mining," *Sustainability*, vol. 11, no. 3, pp. 1-14, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[31] Jafar Alzubi, Anand Nayyar, and Akshi Kumar, "Machine Learning from Theory to Algorithms: An Overview," *Journal of Physics: Conference Series*, vol. 1142, pp. 1-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[32] Majesty Eksa Permana et al., "Sentiment Analysis and Topic Detection of Mobile Banking Application Review," *2020 5th International Conference on Informatics and Computing*, Gorontalo, Indonesia, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[33] Vinod Kumar Chauhan, Kalpana Dahiya, and Anuj Sharma, "Problem Formulations and Solvers in Linear SVM: A Review," *Artificial Intelligence Review*, vol. 52, pp. 803-855, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[34] Sally S. Tinkle et al., "*An Outcome Evaluation of the National Institutes of Health Director's New Innovator Award Program for Fiscal Years 2007-2009*," Science & Technology Policy Institute, 2007. [Google Scholar]
[35] Sri Handika Utami, Anton Ade Purnama, and Achmad Nizar Hidayanto, "Fintech Lending in Indonesia: A Sentiment Analysis, Topic Modelling, and Social Network Analysis Using Twitter Data," *International Journal of Applied Engineering and Technology*, vol. 4, no. 1, pp. 50-56, 2022. [Google Scholar] [Publisher Link]
[36] Mohammed Bahja, "Identifying Patient Experience from Online Resources via Sentiment Analysis and Topic Modelling Approaches," *Thirty Ninth International Conference on Information Systems*, San Francisco, pp. 1-9, 2018. [Google Scholar] [Publisher Link]
[37] Abdullahi Sidow Osman, "Data Mining Techniques: Review," *Data Science and Networking*, vol. 2, no. 1, pp. 1-4, 2019. [Google Scholar] [Publisher Link]
[38] Christoph Schröer, Felix Kruse, and Jorge Marx Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526-534, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[39] Aditya Wiha Pradana, and Mardhiya Hayaty, "The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts," *Kinetik: Game Technology, Information System*, *Computer Network*, *Computing*, *Electronics*, *and Control*, vol. 4, no. 4, pp. 375-380, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[40] Ezz El-Din Hemdan, Marwa A. Shouman, and Mohamed Esmail Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," *arXiv*, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[41] Paul Mooijman et al., "The Effects of Data Balancing Approaches: A Case Study," *Applied Soft Computing*, vol. 132, pp. 1-32, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[42] Der-Jang Chi, and Zong-De Shen, "Using Hybrid Artificial Intelligence and Machine Learning Technologies for Sustainability in Going-Concern Prediction," *Sustainability*, vol. 14, no. 3, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[43] Rindu Hafil Muhammadi, Tri Ginanjar Laksana, and Amalia Beladinna Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," *Khazanah Informatika: Journal of Computer Science and Informatics*, vol. 8, no. 1, pp. 59-71, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[44] W.P. Ramadhan, S.T.M.T. Astri Novianty, and S.T.M.T. Casi Setianingsih, "Sentiment Analysis Using Multinomial Logistic Regression," *2017 International Conference on Control*, *Electronics*, *Renewable Energy and Communications*, Yogyakarta, Indonesia, pp. 46-49, 2017. [CrossRef] [Google Scholar] [Publisher Link]
[45] Elshrif Elmurngi, and Abdelouahed Gherbi, "Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques," *DATA ANALYTICS 2017: The Sixth International Conference on Data Analytics Detecting*, pp. 65-72, 2017. [Google Scholar]
[46] Thanh-Nam Doan, and Tuan-Anh Hoang, "Benchmarking Neural Topic Models: An Empirical Study," *Findings of the Association for Computational Linguistics: ACL-IJCNLP*, pp. 4363-4368, 2021. [CrossRef] [Google Scholar] [Publisher Link]