A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024

Boy Setiawan

Ilmu Komputer, Universitas Nusa Mandiri
boye.setiawan@gmail.com

Abstract— **The landscape of sentiment analysis applications in Indonesia is on the rise with the many published papers on the subject over the years. The need to predict sentiment coincides with the rise of social media and how the public uses it to express sentiments toward an interesting topic. The lack of tools for working with the Indonesian language has brought the invention of libraries to tackle the difficulty and uniqueness of the language on various topics from diverse data sources. The introduction of Sarkawi as a stemmer helps researchers overcome dimensionality problems commonly found with text processing, and boosts the performance of machine learning (ML) models. Using InSet as a lexicon dictionary capable of performing sentiment prediction has started gaining popularity for automatic labeling. The development of IndoBERT, an advanced neural network (NN) large language model (LLM) specifically trained from a large Indonesian text corpus capable of more than sentiment analysis, has gained traction both for automatic labeling and prediction models. Although the majority of research revolves around Naïve Bayes (NB), State Vector Machine (SVM), and K-Nearest Neighbor (KNN) the future of sentiment analysis applications in Indonesia could be heading towards a more advanced deep learning architecture. Finally, this study is intended as a basis for future research in the applications of sentiment analysis in Indonesia and the development of the language.**

***Keywords*— Sentiment Analysis, Naïve Bayes, State Vector Machine, K-Nearest Neighbor, IndoBERT**

I. INTRODUCTION

As the amount of information continues to grow, the need to do a fast, efficient, and reliable sentiment analysis has been rapidly evolving as a solution to extract and evaluate subjective information from textual data. As a field within natural language processing (NLP), sentiment analysis has gained a significant boost over the past decade driven by the considerable volume of user-generated content across social media platforms, blogs, reviews, and other digital channels. The primary objective of sentiment analysis is to determine whether a piece of text expresses a certain sentiment that could be categorized for example positive, negative, or neutral, and can provide valuable insights into public opinion, consumer behavior, and emotional responses to various events, condition or situation [1]. The unique ability to produce an automatic sentiment from text has made sentiment analysis a prime candidate across a wide array of information domains not being limited only to communications or marketing, but applicable to social studies as a tool for understanding consumer behavior or satisfaction but also for policy acceptance in general [2].

As of January 2024, there were 139 million active social media users identified from Indonesia or an equivalent of 49.9% percent of the total population [3]. Around 126.8 million are users aged 18 and above, which account for almost 64.8% of the total population of the same age. This statistic puts Indonesia as one of the highest users of social media in the world, where information and the population's opinions are scattered across social media platforms such as Meta, YouTube, X, TikTok, and Instagram. The huge amount of text that can be exploited to be transformed into something more valuable such as sentimental analysis has driven a lot of studies in this field especially on the behaviors and opinions of social media users in Indonesia since its application is considered faster, feasible, and cost-effective than a manual survey.

The diffusion of ML techniques into sentiment analysis has nothing but changed the field of how we interpret and classify emotion expressed in textual data. The advancement of sophisticated models and enhanced methods have improved the accuracy and efficiency of classifying sentiment tasks [4]. This can be seen from the number of journals on the application of ML to predict sentiment analysis, especially in Indonesia, which is the focus of this study. Although the field of NLP has seen a tremendous amount of research and breakthroughs in the English language, but very little in the domain of another language where the Indonesian language is not an exception. These challenges had not been a barrier to the use and application of ML in sentiment analysis in the Indonesian language but rather increased the research and tools needed to solve it.

The importance of preprocessing text data in the Indonesian language cannot be overstated, as it significantly impacts the accuracy of sentiment classification. The use of Sastrawi as a prominent library that we encountered in our study has shown that a library designed specifically for NLP in the Indonesian language indicates a rise in the sentiment analysis of the language. The library itself which focuses on stemming by reducing words to their base root form, is crucial for various text processing including sentiment analysis, text classification, and information retrieval [5].

The extensive use of X (formerly known as Twitter) where data is open and accessible to be retrieved and analyzed, has emerged as one of the most prevalent social media platforms where sentiment analysis is extensively used to understand public sentiments, especially in Indonesia. Our study found that many applications of ML in sentiment analysis often utilized datasets derived from X besides other dataset such as YouTube or Google Play Store. ML-based approaches particularly supervised techniques, utilize various algorithms such as NB, SVM, Logistic Regression (LR), and Decision Tree (DT) among others to classify sentiments in the Indonesian language. More advanced techniques such as particle swarm optimization (PSO) to produce a better model have also been employed, which enriched the extensive use of ML to predict sentiment. We found that the domain of sentiment analysis extends beyond mere opinion mining, but plays a crucial role in various sectors including politics, marketing, and public health. For example, the previous study by [6] found that sentiment analysis has been employed to predict public reactions to government policies during the COVID-19 pandemic, providing insights into public sentiment and aiding in decision-making processes, which our study found is still being studied more than a year after the global pandemic has lessened down which shows that sentiment analysis does not stop only to understand the present but the past as well.

This survey presents the latest study on sentiment analysis applications in Indonesia using ML both supervised and unsupervised learning problems as well as social network analysis (SNA). Review papers are needed to show an overview tendency and progress highlight any findings and discuss future studies, in the hope the obtained knowledge is useful and can be translated to useful practice and application. There are a few published review papers such as the study from [7] which focused on the use of sentiwordnet for sentiment analysis, but this study was conducted in 2018. However, the use of sentiwordnet could not be found during our study, and the use of Valence Aware Dictionary and Sentiment Reasoner (VADER) is more prevalent by doing a pre-translation prior to using the library. The study by [8] conducted in 2022, did not discuss deep enough the use and application of sentiment analysis in the Indonesian language. Thus, this survey aims to present a more comprehensive summary of the use and application of sentiment analysis in Indonesia and the language including the technique and methods used to accomplish it. The contributions of this review are:

1. An in-depth view of the methods and techniques used for sentiment analysis in Indonesia and the language;
2. A summary of the problem domain in the application of sentiment analysis in Indonesia and the language;
3. List of commonly used sources of dataset;
4. Comparison of ML applied in the studies of sentiment analysis in Indonesia and the language.

We believe that future studies in this topic will benefit greatly from the results and findings of our study, and will help the development and contribution to enhance understanding and knowledge of sentiment analysis, especially in the Indonesian language.

This paper is organized as follows; “Research Method” explains the methodology used in this survey. “Result and Discussion” presents “Problem Domain” as the main idea and necessity of sentiment analysis. “Dataset” shows the common dataset used in the papers. “Labeling” describes the many methods used to give sentiment to a text. “Visualization” explains the common graphical tools used. “Machine Learning” gives a summary of the various ML being applied. “Tools and Languages” describes the technology, methods, and techniques used in sentiment analysis. Finally, we conclude the survey in “Conclusions”.

II. RESEARCH METHOD

This work is based on a literature review of sentiment analysis applications in Indonesia and the language. The main objective is papers published in Indonesian national journals accredited to Sinta or international journals with sentiment analysis on the Indonesian language as the main study. On the papers found, we summarize and calculate the works reported and classify them of interest and findings.

A. Search keywords

We use Google Scholars with the search term “sentimen analisis bahasa indonesia” and set the filter date from 2023-2024 to find the relevant papers. The result given by the query returned a list of journals that were used as the source information for this study.

B. Data Sources

The papers included in this survey were retrieved from various journals indexed in Sinta and other diverse quality databases such as Scopus.

C. Article Inclusion/Exclusion Criteria

We decided which articles are eligible for the survey under the following inclusion/exclusion criteria:

a. Inclusion criteria:
- Manuscripts written in Indonesian or English and published by indexed journals in Sinta to ensure the focus of the study originated from Indonesia.
- Manuscripts written in the English language and published by indexed journals in Scopus that enhanced the advancement of sentiment analysis in the Indonesian language.

b. Exclusion criteria:
- Manuscripts using machine learning marginally or without solid sentiment analysis conclusions
- Manuscripts in preprint without peer review

D. Article Selection

We sorted the papers that were identified by date of publication, of which were filtered by the inclusion/exclusion criteria, and chose 75 articles as the final selection.

III. RESULT AND DISCUSSION

A. Problem Domain

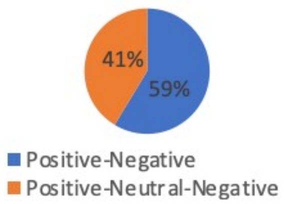

Figure 1. Number of classes being predicted

With most sentiment analysis studies, the goal is to predict two or three sentiments from a text. The most common is to predict whether a text is positive or negative with 59% of journals focused on this issue. this compared to 41% of other journals which focused on predicting a positive, neutral, and negative sentiment from a text as shown in Figure 1. In brief summary, after filtering all the domain problems being studied, we concluded there are 5 major interests; government policy, COVID issues, general elections, national or international events and issues, and commercial products and services.

1) Government Policy

The focus of sentiment analysis on this issue coincides with what happened with Indonesian government policy implementation during 2023-2024. For example, the studies of [9], [10], and [11] focus on the public sentiment about the rising prices of fuel while [12] focuses on the effect of Direct Cash Subsidy (BLT) and how the public perceives this issue on positive or negative sentiment. On the other hand the study of [13] focus on the rising price of cooking oil which is a contradiction with the status of Indonesia as the biggest producer of palm oil in the world.

Although the impact of rising prices could have produced a negative sentiment on public opinion, another study by [14], [15], [16] are doing a different study on predicting more positive sentiment issues of government policy on accelerating the use of electrical vehicles and the impact of speed train development infrastructure between Jakarta and Bandung.

Another major issue is the plan to move the capital from Jakarta to Ibu Kota Nusantara (IKN) which was studied by [17], [18], [19] to see the public sentiment on how the decision of the government to move the capital is considered a positive or negative policy implementation. While the issue of public health quality is the subject study of [20], [21] which also attracts interest from the public perspective like the new omnibus health laws still in drafted and the new health insurance programs of the Health Insurance Administration Body (BPJS). The other major issue is education which has been thoroughly studied by [22], and [23] in their respective papers about the new policy of Merdeka Belajar Kampus Merdeka (MBKM) and how the implementation and the effect of it on universities and their students. The last sensitive government policy is studied by [24] on how the public sentiment toward the recruitment process of contract-based government employees (P3K).

The abundance of papers on sentiment analysis which focuses on how the public opinions towards government policy is quite significant, attracted the respective authors of the papers to conduct and to provide a fast answer and a degree of certainty to understand how the public perceives the policy.

2) Covid Issues

The year 2019-2021 has been marked by a worldwide pandemic caused by the virus known as Covid-19. During this pandemic the world was at a standstill, a lot of public contact is prohibited and laws were enacted to help prevent and reduce the widespread and deathly disease. But the world has recovered since then and has started to forget the terrifying pandemic behind it. However, it is surprising that after 2 years, a number of studies regarding sentiment analysis are still being conducted, especially on certain issues that occurred during the pandemic period.

The main question regarding the pandemic was about vaccination which is the main study of [24], [25], [26], [27], [28] on their respective papers. The studies were done to find out the public sentiment towards the quality and safety of the vaccines. Despite the assurance of the government, a rejection or denial about the necessity and safety of the vaccines were still found. This objection was mostly caused by the number of hoaxes that spread rapidly during the pandemic since the need to find any information about the disease online was quite massive.

The effect of hoaxes after the pandemic has also attracted a study by [30] which differs from other studies about COVID-19, where the authors did not rely on ML to predict sentiment but rather utilized other NLP text mining techniques to find words associated with hoax information and derived understanding from it.

The number of papers on the COVID-19 pandemic period which is still being studied after it was declared ended, shows that past event sentiment can be used to know what happened to derive a better understanding of why it happened.

3) General Elections

The general elections of 2024 have also been studied thoroughly prior to the event taking place, which mainly focused on the presidential election. To find out the public sentiment about this issue, studies by [31], [32], [33], and [34] were done by the respective authors to see the effects of the election on the public sentiment. A different study by [35] focuses on the public sentiment towards the implementation of e-voting which is a new idea in the general elections. The study by [36] focused on the indication of election fraud and how the public sentiment perceives how will this issue impact the general election.

The different studies on this issue show how sentiment analysis can be used to perceive how the general public sentiment towards a nation-wide process, and how sentiment analysis can reduce uncertainty.

4) National Or International Events and Issues

There are two global events that attract a study about how the Indonesian public sentiment opinions on the subject. One study by [37] focuses on the global economic recession of 2023 and how the public sentiment towards the impact will be on their economic situation and spending power. Another study by [38] discusses the immigration of the Rohingya refugees on the sea of Aceh and how the public sentiment responds to this issue. One notable different focus of the study is the sentiment of the American veterans towards the decision to leave Afghanistan [39] which differs from other papers on the direction of the sentiment being studied.

On a national scale, three issues attract national interest during our study time period. The issue of fintech-based debt, which is the focus study of [40], [41]. This issue has taken the interest of the public driven by the amount of bad news and negative reviews. Despite the negative precedents, both studies found there were also positive and neutral sentiments on social media about this issue. The second issue that attracts national public interest is the topic of premarital sex which is considered immoral, but the latest affairs suggest that this conduct is quite rampant among young people. This issue is the subject study of [42] in which the authors found an abundant negative sentiment, there was also a neutral one.

The last issue is about khilafah which is a political institution based on Islamic law. This issue has always been perceived and associated with radicalism and is the main study of [43] to determine what the public sentiments are towards this misinterpretation. A different approach used to predict sentiment was conducted by [44] on the tragic event of Kanjuruhan which killed 135 football supporters, to see if the concern for raised football match safety is a non-orchestrated and naturally flowing response to the tragedy.

The various issues being studied by the respective authors on this subject by applying sentiment analysis show that sentiment on national or international events and issues can be predicted to answer the curiosity of the general public.

5) Commercial Product/Services

The types of problems in this subject are diverse and range from software, to products, and services.

5.1. Software

Papers from [45] focused their study on how the public sentiment views online games, where quarantine during the COVID-19 pandemic many people were heavily engaged. On the other hand, a study from [46] was conducted to compare which is better between a particular game but played on a different media in this case mobile phone. A different study from [47] uses the same question domain on crowdfunding applications to see which one has the highest positive sentiment.

As the public starts to embrace online entertainment like online games, the enjoyment of streaming content has also been the focus of study by [48] to find out the public perception of their application, service, and media content. The prominent rise of TikTok as a new social media platform has also been studied by [49] along with the application for making and editing video content for TikTok by [50] to perceive how the application performance is on the public view.

5.2. Product

A study by [51] asks question on the public sentiment of the recent popularity of skin care products in Indonesia, while a study by [52] asks about the public sentiment about a networking product. A comparison study between E-commerce platforms in Indonesia is the focus study of [53] while the quality of products from online shopping Lazada is the focus study of [54]. A different study by [55] asks about public sentiment on the prohibition of cough syrup by the government. The release of ChatGPT in late 2022 as a groundbreaking and innovative chatbot that could help its users with many multitasking tasks, has also attracted studies on the positive and negative response of the public from its use, such as the study from [56], [57], [58]. The many options of online learning platforms that are available today have also been studied by [59] to understand the public sentiment towards materials quality.

5.3. Services

The banking system services have been studied by the works of [60], and [61] on how their services is perceived by their customers. In another study, the quality of services provided by major telecommunication providers is the main study subject of [62], [63], [64]. In the study of [65], the author's focus is on the service provided by the mobile version of the export tax application and the sentiments towards it.

The number of papers published on this subject shows that the application of sentiment analysis has been widely used to predict the public perception of various fields of commercial products and services.

B. Dataset

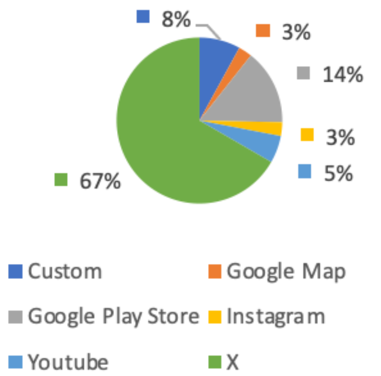

Figure 2. Sentiment analysis dataset source

As shown in Figure 2, the main source of datasets for sentiment analysis comes from X. There are two main ways of acquiring the dataset in general, the first is by accessing an API from the provider. The other one is by scrapping the information needed from the web page directly. From the two, X provides an easy access API which is the main reason why it is popular among researchers. A few authors rely on scrapping the dataset from X such as [66] who uses a library to gather data, but most authors use API to gather data from X.

The X Open Data Platform serves as a critical resource for researchers and practitioners across various fields, enabling the collection and analysis of vast amounts of user-generated content without the constriction of any certain language helps boost its popularity in sentiment analysis. Besides the easiness of collecting the data by accessing its API and the number of data given which usually suffice for study and research purposes without the need to purchase it until a certain amount of data. It also helps that many sentiment analysis tools provide an easy-to-use library or module to help researchers obtain the data with the use of API keys freely provided by X. Another social media platform is Instagram but as reported in their papers, the respective authors rely on scrapping rather than using the social media platform API.

In our study, the most scrapped dataset source comes from Google products. Websites such as Google Play Store, YouTube, and Google Maps is the second source of datasets for sentiment analysis in Indonesia. Most problem domains from these datasets are to analyze reviews or comments on the website about a certain application or site. One notable research is to perceive public sentiment about a tourist attraction site by using Google Maps reviews of visitors to the site [67]. The millions of applications available in the Google Play Store are also a source of opportunity to research public sentiment about applications especially one with the same features as the others. An example of this is the study by [47], [53], [59] where comparisons between similar applications were conducted to perceive the sentiments towards a certain application and to determine the best application according to the reviews and comments. The many news channels on YouTube where many comments are being posted for certain news topics is the subject study of [39], [68].

A study by [54] used a dataset available at Kaggle about the public sentiment towards a certain online shopping platform, which compared to the number of rows being studied outnumbered other studies by 10-30 times on average. One notable study used various types of datasets including Twitter, Instagram, and Websites done by [69] with the main focus not only to predict sentiment analysis on certain domain problems or issues but also to build a model that could be used as a general model to predict sentiment in the Indonesian language. In their study, the authors used multiple topics with many datasets originating from X.

Table 1. Datasets used by [69] reproduced from the papers

No	Topic	Platform Source	#Samples
1	Cyberbullying	Instagram	400
2	Indonesian 2013 curriculum	X	710
3	TV talk show	X	400
4	Cellular service provider	X	300
5	Film review	X	200
6	Head region election	X	900
7	Data breach incident	X	1,060
8	Others	X	10,805
9	Tourist attraction Travel	Website	200

C. Labeling

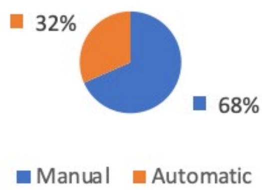

Figure 3. Dataset labelling

Sentiment analysis as a classification and supervised learning needs labeled training data, which is essential for the ML models to train on. The problem of labeling in sentiment analysis comes down to the fact that language is very subjective. Unlike numbers or any other quantitative values that can be derived from a calculation, text-based data can be interpreted differently according to its context. As shown in Figure 3, it is understandable that manual labeling is the dominant way used to label the dataset by relying on an expert opinion on the text in question. This approach is also difficult, especially when the number of data increases or the complexity of the text varies. Text from social media is a prime example, because of the limitation of characters that the platform allows inspired the users to invent and use ingenuity to express their sentiment with emojis, etc. The use of emojis could be an advantage for sentiment analysis, on the other hand, the combination of them makes a text even more subjective and affects the analysis. In our study, we conclude that there are two ways of manual labeling being used; self-labeling by the paper's authors and using a language expert opinion.

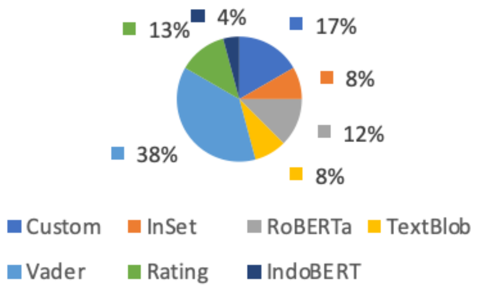

Figure 4. Automatic labeling tools

To overcome the difficulty of manual labeling, some studies rely on automatic labeling to speed up the process. Figure 4 depicts the options on automatic labeling as we discovered in our study. For datasets originating from the Google Play Store, the use of ratings can be used to determine sentiment by implementing a simple calculation such as the study of [45], [49], [50]. But others rely on the use of a lexicon library or a pre-trained LLM which will calculate the text given into a prediction of the sentiment. Below is a list of methods used for automatic labeling we found in our study:

1. Vader

Vader is a specialized tool for sentiment analysis, particularly effective in analyzing social media text. It utilizes a lexicon-based approach that is finely tuned to the nuances of informal language, such as slang, emoticons, and various punctuation styles, which are prevalent in platforms like X. This optimization allows Vader to accurately assess sentiment by calculating a composite score that categorizes text into positive, negative, or neutral sentiments based on the intensity of the emotional content expressed [70].

As this study focuses on sentiment analysis in the Indonesian language, the use of Vader as a tool for automatic labeling is surprising since the library is exclusively used only for English. This means that the text being predicted must be translated into English prior to the automatic sentiment labeling. We found only the study of [39] that works with English text and no translation is needed, other studies using the Vader library make use of a translation library such as Google Translation.

2. Robustly optimized BERT approach (RoBERTa)

RoBERTa is an advanced NLP model developed by Facebook AI that was built on the original Bidirectional Encoder Representations from Transformers (BERT) model with more improvements. These modifications have made RoBERTa one of the leading models in various NLP benchmarks, excelling in tasks such as text classification, sentiment analysis, and question-answering. The advantage of using RoBERTa for automatic labeling is the ability to use the model to recognize sentiment from the Indonesian language directly. This automatic language recognition ability is used by the study of [15] to perceive the public sentiment for electric vehicles in the Indonesian language.

3. Indonesia Sentiment Lexicon (InSet)

InSet is an Indonesian sentiment lexicon built to recognize opinion in text, predict it into either positive or negative, and could be utilized to analyze public sentiment towards a particular topic, event, or product. Composed using a collection of words from tweets, this library was constructed by manually weighting each word and adding enhancement by stemming and synonym set [71]. A study by [47] reported using InSet to decide which application has the best review between two crowdfunding applications using Google Play Store reviews.

Since the library is built from the bottom up as a lexicon library for the Indonesian language, one study from [38] conducted research to compare the performance of automatic labeling classification between Vader and InSet. The study found that an SVM performs better with an InSet label than Vader for sentiment analysis using X and that InSet is better at recognizing negative sentiment compared to Vader. In the end, the respective authors conclude that manual labeling is needed to compare the results between the two libraries to have a definitive conclusion.

4. TextBlob

TextBlob is a Python library that provides a simple API for common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, and more. The easy integration with Python as a de-facto programming language of data science makes these libraries easy to understand and implement. TextBlob only recognizes the English language just like Vader, so require a translation pre-processing. One study that uses this library without a translation pre-processing is by [26] which focuses on the public sentiment toward COVID-19 vaccines that the World Health Organization (WHO) approved.

5. IndoBERT

IndoBERT is a variant of the BERT model, which has been meticulously trained on an extensive dataset comprising over 220 million words, sourced from various Indonesian platforms, including Indonesian Wikipedia, news articles from Kompas, Tempo, Liputan6, and Korpus Web Indonesia [72]. Specifically designed to handle tasks in the Indonesian language, IndoBERT excels in predicting sentiment in the Indonesian language. A study by [14] uses IndoBERT as an automatic labeling to predict public sentiment for electric vehicles among X users in Indonesia.

6. Custom Dictionary

In our research, we found there are also studies by [10], [13], and [31] that use a custom lexicon dictionary consisting of two dictionaries. One is a dictionary containing positive words in Indonesia and the other is the negative words. The final result as stated in the papers was that two studies concluded an accuracy lower than 80%, and there is only one paper that reported a 90% accuracy performance. Both of the results cannot be verified since there is no example of the dictionary being used in the study to compare with.

D. Visualization

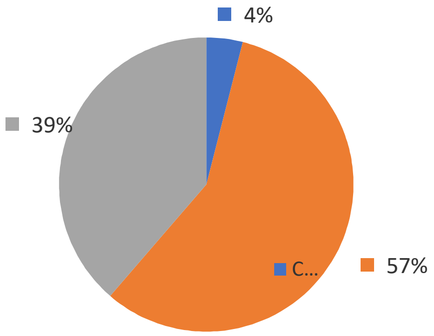

Figure 5. Visualization preferences

The use of visualization to help understand the sentiment being analyzed helps to give a bird's eye view as a whole and the textual cues of the text. This tool is not only nice to look at, but also expresses an emotional landscape of the text. In some cases, the use of visualization can enhance and emphasize the most dominant expression in the text. as shown in Figure 5, word clouds are the most common visualization used in our study. This is not surprising since word clouds are used to highlight the most frequent words in the text being analyzed.

The use of these types of visualization depends on the final end result of the study. Word clouds are used by papers with end-result studies of automatic sentiment prediction only, while social network analysis (SNA) uses word clouds and cluster visualization. An example of SNA to predict sentiment analysis is the studies done by [30], [44] in which a social event that captivates the public attention is being studied to determine the public sentiment and to answer whether a viral online topic is naturally spreading or orchestrated by a certain social media account. The other notable difference of the SNA study is the use of several types of visualization such as hierarchical dendrogram which depicts a tree-like diagram that shows the arrangement of clusters produced by hierarchical clustering, and how data points or groups are related to one another based on their similarity or distance [30]. Another type of visualization is the Fruchterman-Reingold layout, an algorithm used for graph drawing, and visualizing networks in a way that makes the structure and relationships within the data clearer [30]. The other one is the layout circle that visualizes nodes arranged in a circular formation, placed at an equal distance from the center of the circle, to see all nodes and their connections in an easy manner [30].

E. Machine Learning

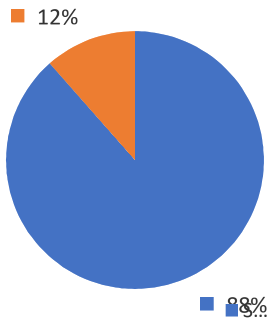

Figure 6. Types of models used for sentiment analysis

As shown in Figure 6, the most dominant model being used for sentiment analysis is shallow learning. The simplicity and easy-to-understand relationships between input features and output give it an advantage over more complex and advanced deep learning model which could involve multiple layers in their architecture. The amount of data to train could be another reason why it has become a common model to choose over a deep learning model which needs a lot of training data. The training time is also much faster and fewer computing resources are needed to train them, especially with a Transformers model that can consume a lot of power, expensive and dedicated hardware such as a graphic processing unit (GPU) to speed up the training time, which is beyond the reach for small research to start with.

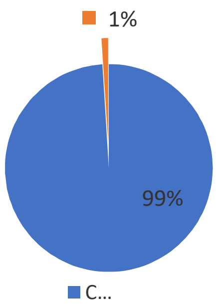

Figure 7. Use cases of shallow models

The majority of shallow models being used are classification type, with one notable difference as shown in Figure 7, where the study of [73] on analyzing the sentiment of online shopping customers uses a clustering approach utilizing K-Means as the model.

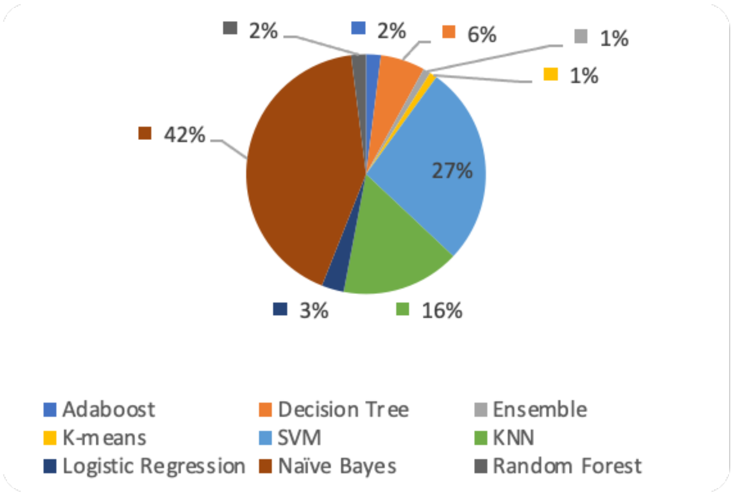

Figure 8. Types of shallow models used for sentiment analysis

For the classification models, NB is the majority followed by SVM and KNN which total 85% of all reported models as shown in Figure 8. While the ensemble model is being used in the study of [39] to predict the sentiment of American veterans on the takeover of Afghanistan by the Taliban, incorporate LR, DT, SVM, and NB in the ensemble model.

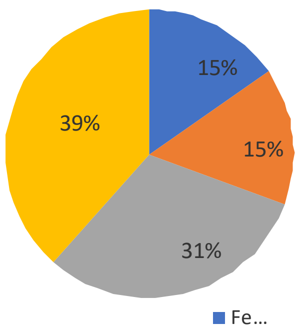

Figure 9. Types of deep models used for sentiment analysis

There are four types of deep learning architecture being used for sentiment analysis as shown in Figure 9. The most common attention-based architecture is IndoBERT such as in the study by [14], [34], [72] which uses transfer learning on the pre-trained IndoBERT model and trains it on the new sentiment dataset. This helps the training process since IndoBERT has already been trained on a huge amount of data, and this method is more accessible to use on an LLM for sentiment analysis.

The other common model is deep sequence architecture that implements gated recurrent unit (GRU) and long-short term memory (LSTM) like the study by [27], [72], [74] which utilize GRU or LSTM on social media data. The use of 1-dimensional convolution to predict sentiment is reported by the study of [74] and in a comparison with a GRU architecture. Another study reported by [35] compares the performance of NB and feed-forward NN to analyze social media data on the topic of e-voting implementation in Indonesia.

Table 2. Models performance

Ref	Model	Accuracy	Dataset	Topic
[9]	KNN	0.94	Twitter	Government Policy
[10]	NB	0.90	Twitter	Government Policy
	SVM	0.78		
[11]	NB	0.68	Twitter	Government Policy
[12]	SVM	0.86	Instagram	Government Policy
[13]	KNN	0.73	Twitter	National or International Events
	Random Forest	0.73		
[14]	IndoBERT	0.99	Twitter	Government Policy
[15]	LSTM	0.63	YouTube	Government Policy
[16]	KNN	1.00	Twitter	Government Policy
[17]	NB	0.87	Twitter	Government Policy
	SVM	0.91		
[18]	NB	0.63	Instagram	Government Policy
	KNN	0.69		
[19]	KNN	0.84	Twitter	Government Policy
[20]	KNN	0.85	Twitter	Government Policy
[21]	SVM	0.98	Twitter	Government Policy
	NB	0.99		
	Random Forest	0.99		
[22]	NB	0.81	Twitter	Government Policy
	DT	0.68		
[23]	NB	0.99	Twitter	Government Policy
[24]	NB	0.96	Twitter	Government Policy
	SVM	0.95		
[25]	NB	0.83	Twitter	Covid-19
	DT	0.78		
[26]	KNN	-	Twitter	Covid-19
[27]	NB	0.85	Twitter	Covid-19
	SVM	0.85		
	LSTM	0.83		
[28]	NB	0.93	Twitter	Covid-19
[29]	SVM	0.81	Twitter	Covid-19
[30]	SNA	-	Custom	Covid-19
[31]	NB	0.63	Twitter	General Elections
	SVM	0.70		
[32]	LR	0.86	Custom	General Elections
[33]	SVM	0.91	Twitter	General Elections
[34]	IndoBERT	0.84	Twitter	General Elections
	NB	0.81		
	SVM	0.80		
[35]	NB	-	Twitter	General Elections
	NN	-		
[36]	NB	0.84	Twitter	General Elections
[37]	NB	0.73	Twitter	National or International Events
	SVM	0.80		
[38]	Lexicon	-	Twitter	National or International Events
[39]	Ensemble	0.75	YouTube	National or International Events
	KNN	0.43		
	DT	0.68		
	SVM	0.65		
	LR	0.75		
[40]	NB	0.80	Twitter	National or International Events
[41]	SVM	0.99	Twitter	National or International Events
	DT	0.97		
[42]	SNA	-	Twitter	National or International Events
[43]	KNN	0.92	Twitter	National or International Events
[44]	SNA	-	Twitter	National or International Events
[45]	NB	0.87	Google Play Store	Commercial Products/Services
[46]	NB	0.96	Twitter	Commercial Products/Services
[47]	Lexicon InSet	-	Google Play Store	Commercial Products/Services
[48]	NB	0.64	Google Play Store	Commercial Products/Services
[49]	SVM	0.84	Google Play Store	Commercial Products/Services
	NB	0.79		
[50]	NB	0.81	Google Play Store	Commercial Products/Services
	SVM	0.86		
[51]	NB	0.86	Twitter	Commercial Products/Services
[52]	KNN	0.83	Custom	Commercial Products/Services
[53]	SNA	-	Twitter	Commercial Products/Services
[54]	NB	0.93	Custom	Commercial Products/Services
[55]	NB	0.65	YouTube	Government Policy
[56]	DT	0.77	Twitter	Commercial Products/Services
	NB	0.77		
[57]	KNN	0.85	Twitter	Commercial Products/Services
[58]	SVM	0.59	Twitter	Commercial Products/Services
	NB	0.47		
[59]	SVM	0.92	Google Play Store	Commercial Products/Services
[60]	SVM	0.88	Twitter	Commercial Products/Services
	NB	0.76		
	LR	0.86		
[61]	NB	0.86	Twitter	Commercial Products/Services
	AdaBoost-NB	0.87		

[62]	NB	0.96	Google Play Store	Commercial Products/Services
[63]	NB	0.85	Twitter	Commercial Products/Services
[64]	SVM	0.80	Twitter	Commercial Products/Services
	SVM+Adaboost	0.80		
[65]	NB	0.80	Google Play Store	Commercial Products/Services
	KNN	0.78		
[66]	KNN	0.79	Twitter	Covid-19
	SVM	0.54		
[67]	SVM	0.93	Google Map	National or International Events
[68]	SVM	0.85	YouTube	General Elections
[69]	Hybrid Model	0.85	Various	Multi-Purpose
[72]	IndoBERT	0.96	Google Play Store	Commercial Products/Services
[73]	K-Means	-	Custom	Commercial Products/Services
[74]	CNN	0.91	Twitter	Government Policy
	GRU	0.91		
	GRU-CNN	0.91		
	CNN-GRU	0.91		
[75]	NB	0.79	Twitter	Government Policy
	SVM	0.85		
[76]	SVM	0.84	Twitter	General Elections
	NB	0.77		
	LR	0.84		
[77]	NB	0.75	Twitter	Covid-19
	KNN	0.49		
[78]	NB	0.87	Twitter	Covid-19
[79]	KNN	0.75	Google Play Store	Commercial Products/Services
[80]	NB	0.90	Twitter	Covid-19
[81]	DT	0.66	Twitter	National or International Events
[82]	SVM	0.72	Twitter	Covid-19
[83]	NB	0.90	Google Map	Commercial Products/Services
	SVM	0.92		
[84]	KNN	0.91	Twitter	General Elections
[85]	NB	0.69	Custom	National or International Events
	KNN	0.85		

Table 2 shows the performance results as reported in the papers by their respective authors. The accuracy results cannot be used to justify whether one model is better than the other since there are differences in methodology, dataset, research goals, and hyperparameters setting that vary and must be considered and verified accordingly.

F. Tools and Languages

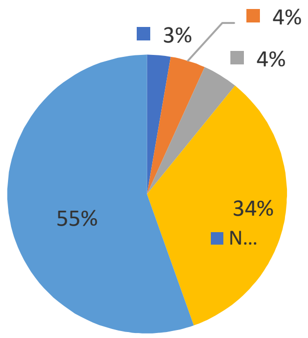

Figure 10. Tools and languages used for sentiment analysis

As shown in Figure 10, we identified not only programming languages but tools used for sentiment analysis as a way to know exactly what is the preference for doing sentiment analysis. The most common language is Python as the de-facto standard language for data science, followed by Rapid Miner a drag-and-drop Java-based application for doing data analytics. R with Rstudio and Orange only took a small portion of the survey, while Netlytic is mainly used for doing SNA.

IV. CONCLUSION

There have been many studies performed in Indonesia using the Indonesian language to predict or analyze sentiment. These studies are conducted commonly with online data especially on social media since this method is easier and more accessible than a survey. The increased attention and the need to perform a fast and accurate sentiment analysis has brought forward tools built especially to tackle the difficulty of the Indonesian language context and syntax. A library like Sastrawi often used as a stemmer, gives an impact by reducing the dimensionality of processing text data by transforming the word to its base root word. This tool can help a lot of studies beyond sentiment analysis because it is applicable to all text processing in the Indonesian language. The use of stop words dictionary commonly used to erase non-important words in a text has also caught our attention during our study. Although well know library provides a stop-word dictionary in the Indonesian language such as the spacy library, we found some studies use a custom stop-word dictionary.

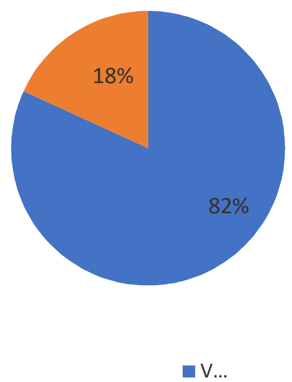

Figure 11. Papers using Vader and InSet

If text mining is not already a daunting task, the various ways people express their sentiments on social media have put another strain by way of emojis and slang words. Fortunately, there is an abundance of slang word dictionaries that are available to download, use, and modify according to the need. The creation of InSet by [71] as a lexicon library to predict sentiment in the Indonesian language could let the researchers focus more on the problem at hand. A study to compare both dictionaries is done by [47] and reported that InSet performs better than Vader. As shown in Figure 11 the use of InSet is still in its infancy, but in the future could surpass Vader or TextBlob because a translation from Indonesian to English is no longer needed to perform automatic labeling,

As shown in Figure 6, most ML used is shallow learning and NB reigns supreme. Together with SVM and KNN, they comprise almost 85% of the total shallow learning used in the application of sentiment analysis. It proves that the majority of research in sentiment analysis revolves around the three ML models, and they will continue to be used in the future.

The use of more advanced and complicated models has started to increase which could produce better results in the study of sentiment analysis. The application of attention-based architecture like Transformers compared to sequence-based architecture like LSTM and GRU is a sign that the traction of the application on sentiment analysis will likely produce a model which understands the context and the position of the text better than the result of a shallow model which only accounts the value of the word. The introduction of IndoBERT, an LLM specifically trained and capable of performing more than sentiment analysis in the Indonesian language has started to make ground for a reliable alternative in sentiment analysis prediction or automatic labeling. This model can also be the best alternative option to the more established lexicon libraries such as Vader, TextBlob, or even InSet because it has been trained on a huge amount of text and sentiment in the Indonesian language. Some studies have utilized this ability to provide automatic labeling such as the study by [14], while other studies utilize the ability of transfer learning and retrain the model for a particular problem such as the study by [34], [72].

It is very clear that at this moment, we are still at an early stage of sentiment analysis but the future is hopeful. The application of sentiment analysis lies beyond a single domain problem but is far stretched to almost anything that can attract the public curiosity about certain events, issues, products, and services. As discussed above, the amount of data is growing every day and the public sentiment curiosity for an answer will never stop and more research on this subject will turn out to answer it.

In this research, we exhaustively reviewed studies on the application of sentiment analysis in Indonesia and the Indonesian language. In our opinion, there will be more tools and libraries to help researchers specifically in the Indonesian language as the interest and the challenge grows. More complex and advanced models will be utilized to give a better result than the models used today because more accurate prediction is needed. The way to give a better prediction is to understand more about the context of the text in question, which proves difficult using the current majority model at the moment.

Finally, we hope that this review will serve as a starting point for other researchers in the application of sentiment analysis, as well as other researchers who are more focused on improving and inventing ways to enhance sentiment analysis in Indonesia. More research, tools, methods, and solutions are required to improve this field.

REFERENCES

[1] A.G. Chifu and S. Fournier, "Sentiment Difficulty in Aspect-Based Sentiment Analysis," *Mathematics*, vol. 11, no. 22, 2023, doi: 10.3390/math11224647.

[2] P. Sánchez-Núñez, C. de las Heras-Pedrosa, and J. I. Peláez, "Opinion Mining and Sentiment Analysis in Marketing Communications: A Science Mapping Analysis in Web of Science (1998–2018)," *Soc. Sci.*, vol. 9, no. 3, 2020, doi: 10.3390/socsci9030023.

[3] S. Kemp, "DIGITAL 2024: INDONESIA." [Online]. Available: https://datareportal.com/reports/digital-2024-indonesia

[4] S. Redjeki and W. Setyawan, "Comparison of Seven Machine Learning Algorithms in the Classification of Public Opinion," *J. TECH-E*, vol. 5, no. 2, 2022, doi: 10.31253/te.v5i1.1046.

[5] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, p. 012017, Jun. 2020, doi: 10.1088/1757-899X/874/1/012017.

[6] B. S. Rintyarna *et al.*, "Modelling Service Quality of Internet Service Providers during COVID-19: The Customer Perspective Based on Twitter Dataset," *Informatics*, vol. 9, no. 1, 2022, doi: 10.3390/informatics9010011.

[7] Sherly Christina and Deddy Ronaldo, "A Survey of Sentiment Analysis Using Sentiwordnet on Bahasa Indonesia," *J. Teknol. Inf.*, vol. 12, no. 2, pp. 69–73, 2018.

[8] T. Walasary, "Survey Paper Tentang Analisis Sentimen," *J. Konstelasi*, vol. 1, 2022, doi: 10.24002/konstelasi.v2i1.5378.

[9] A. J. Arifin and A. Nugroho, "Uji Akurasi Penggunaan Metode KNN dalam Analisis Sentimen Kenaikan Harga BBM pada Media Twitter," *Progresif J. Ilm. Komput.*, vol. 19, no. 2, pp. 700–709, 2023.

[10] Ferdi and V. Ayumi, "Analisa Sentimen Mengenai Kenaikan Harga BBM Menggunakan Metode Naïve Bayes dan Support Vector Machine," *JSAI J. Sci. Appl. Inform.*, vol. 6, pp. 1–10, Feb. 2023, doi: 10.36085/jsai.v6i1.4628.

[11] A. Abdillah M. ,. Fikry, M. ,. Nazir and Insani, "Analisa Sentimen Terhadap Kenaikan BBM di Twitter (X) Menggunakan Naive Bayes Classifier," *J. CoSciTech Comput. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 65–74, 2024.

[12] R. Salam, M. Jamil, Y. Ibrahim, R. Rahmaddeni, S. Soni, and H. Herianto, "Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine: Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, pp. 27–35, May 2023, doi: 10.57152/malcom.v3i1.590.

[13] C. Yanti, N. Agustini, N. Ginantra, and D. Wulandari, "Perbandingan Metode K-NN Dan Metode Random Forest Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia," *J. MEDIA Inform. BUDIDARMA*, vol. 7, p. 756, Apr. 2023, doi: 10.30865/mib.v7i2.5900.

[14] R. Merdiansah, S. Siska, and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. Dan Sist. Inf. JIKOMSI*, vol. 7, no. 1, pp. 221–228, Mar. 2024, doi: 10.55338/jikomsi.v7i1.2895.

[15] A. S. Widagdo, K. N. Qodri, F. E. N. Saputro, N. A. R. Putri, and others, "Analisis Sentimen Mobil Listrik di Indonesia Menggunakan Long-Short Term Memory (LSTM)," *J. FASILKOM*, vol. 13, no. 3, pp. 416–423, 2023.

[16] Z. R. Hakim and S. Sugiyono, "Analisa Sentimen Terhadap Kereta Cepat Jakarta–Bandung Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," *J. Sains Dan Teknol.*, vol. 5, no. 3, pp. 939–945, 2024.

[17] A. M. Siregar, "Analisis Sentimen Pindah Ibu Kota Negara (IKN) Baru pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *Fakt. Exacta*, vol. 16, no. 3, 2023.

[18] D. Pramana, M. Afdal, M. Mustakim, and I. Permana, "Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neightbors," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1306–1314, 2023.

[19] C. Huda and M. Betty Yel, "Analisa Sentimen Tentang Ibu Kota Nusantara (IKN) Dengan Menggunakan Algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes," *J. Ilmu Komput. Dan Sist. Inf. JIKOMSI*, vol. 7, no. 1, pp. 126–130, Feb. 2024, doi: 10.55338/jikomsi.v7i1.2846.

[20] C. Tupari; Abdullah, Syaukani; Chairani, "Visualisasi Data Analisa Sentimen RUU Omnibus Law Kesehatan Menggunakan KNN dengan Software RapidMiner," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 261–268, 2023.

[21] Z. Ardika and A. D. Wowor, "Analisis Sentimen Masyarakat Terhadap Program Badan Penyelenggara Jaminan Sosial (BPJS) Menggunakan Data Twitter," *JIPI J. Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 90–99, 2024.

[22] N. Amalia, T. Suprapti, and G. Dwilestari, "Analisis Sentimen Pengguna Twitter Terhadap Pelaksanaan Kurikulum MBKM," *E-Link J. Tek. Elektro Dan Inform.*, vol. 18, no. 1, pp. 57–64, 2023.

[23] J. T. Kumalasari and A. Merdekawati, "Analisis Sentimen Terhadap Program Kampus Merdeka Pada Twitter Menggunakan Metode Naïve Bayes, Union dan Synthetic Minority Over Sampling Technique (SMOTE)," *Satin-Sains Dan Teknol. Inf.*, vol. 9, no. 1, pp. 01–12, 2023.

[24] F. N. Hidayat and S. Sugiyono, "Analisis Sentimen Masyarakat Terhadap Perekrutan PPPK Pada Twitter Dengan Metode Naive Bayes dan Support Vector Machine," *J. Sains Dan Teknol.*, vol. 5, no. 2, pp. 665–672, Dec. 2023, doi: 10.55338/saintek.v5i2.1359.

[25] A. Z. Ahmad, E. Asril, M. Sadar, Y. Turnandes, and others, "Analisis Sentimen Opini Terhadap Vaksin Covid-19 Pada Media Sosial Twitter Menggunakan Naïve Bayes dan Decision Tree," *ZONAsi J. Sist. Inf.*, vol. 5, no. 1, pp. 100–110, 2023.

[26] P. Arsi, I. Prayoga, and M. Asyari, "Klasifikasi Sentimen Publik Terhadap Jenis Vaksin Covid-19 yang Tersertifikasi WHO Berbasis NLP dan KNN," *J. MEDIA Inform. BUDIDARMA*, vol. 7, p. 260, Jan. 2023, doi: 10.30865/mib.v7i1.5418.

[27] Kristiyanti, D. A. and Hardani, Sri "Sentiment Analysis of Public Acceptance of Covid-19 Vaccines Types in Indonesia using Naïve Bayes, Support Vector Machine, and Long Short-Term Memory (LSTM)," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 7, no. 3, Jun. 2023, doi: 10.29207/resti.v7i3.4737.

[28] M. Ridho, A. M. Husein, V. B. Halawa, N. A. Pasaribu, S. Kumar, and others, "Sentiment Analysis of Indonesia Covid-19 Vaccine on Twitter Using Naïve Bayes Classifier," *Data Sci. Indones. DSI*, vol. 3, no. 2, pp. 90–97, 2023.

[29] S. Sulastri and F. A. Nur, "Analisa Sentimen Twitter Vaksin Covid-19 di Indonesia dengan Metode Support Vector Machine," *Kesatria J. Penerapan Sist. Inf. Komput. Dan Manaj.*, vol. 5, no. 3, pp. 1244–1252, 2024.

[30] A. Purnajaya and Y. Pernando, "Analisa Sentimen Informasi Hoaks Pasca Pandemi Covid-19 dengan Text Mining," *J. Comput. Syst. Inform. JoSYC*, vol. 4, pp. 1–10, May 2023, doi: 10.47065/josyc.v4i3.3358.

[31] U. R. H. Baba, "Analisa Sentimen Menjelang Pemilihan Umum Presideen 2024 di Indonesia Menggunakan Perbandingan Performa Support Vector Machine (SVM) dan Naïve Bayes," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 3, pp. 11972–11990, 2024.

[32] A. Hidayati and A. Fitriani, *Analysis of 2019 Election Sentiment in Online News Titles Using the Logistic Regression Method: Analisa Sentimen Pemilu 2019 pada Judul Berita Online Menggunakan Metode Logistic Regression*. 2023. doi: 10.21070/ups.493.

[33] L. Damayanti and K. Lhaksmana, "Sentiment Analysis of the 2024 Indonesia Presidential Election on Twitter," *Sinkron*, vol. 8, pp. 938–946, Mar. 2024, doi: 10.33395/sinkron.v8i2.13379.

[34] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using Bert Language Models," *J. Ilm. Tek. Elektro Komput. Dan Inform.*, vol. 9, no. 3, pp. 746–757, Aug. 2023, doi: 10.26555/jiteki.v9i3.26490.

[35] M. R. Adipratama and N. Safriadi, "Analisis Sentimen Terhadap Rencana Penerapan E-Voting Pada Pemilu di Indonesia," *J. Linguist. Komputasional*, vol. 7, no. 1, pp. 26–30, Mar. 2024, doi: 10.26418/jlk.v7i1.214.

[36] D. S. Nugroho, I. F. Hanif, M. A. Hasbi, F. Fredianto, A. M. Saputra, and R. Zildjian, "Analisis Sentimen Dugaan Pelanggaran Pemilu 2024 Berdasarkan Tweet Menggunakan Algoritma Naïve Bayes Classifier: Sentiment Analysis of Alleged 2024 Election Fraud Based on Tweets Using the Naïve Bayes Classifier Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1169–1176, 2024.

[37] S. A. Sutresno, "Analisis Sentimen Masyarakat Indonesia Terhadap Dampak Penurunan Global Sebagai Akibat Resesi di Twitter," *Build. Inform. Technol. Sci. BITS*, vol. 4, no. 4, pp. 1959–1966, 2023.

[38] Muhammad Fernanda Naufal Fathoni, Eva Yulia Puspaningrum, and Andreas Nugroho Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Modem J. Inform. Dan Sains Teknol.*, vol. 2, no. 3, pp. 62–76, Jul. 2024, doi: 10.62951/modem.v2i3.112.

[39] H. Leidiyana, "Ensemble Stacking Dalam Analisa Sentimen Reaksi Veteran Militer AS Terhadap

Pengambilalihan Afghanistan Oleh Taliban," *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 23–28, 2023.

[40] M. I. Ghozali, W. H. Sugiharto, and A. F. Iskandar, "Analisis Sentimen Pinjaman Online di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK Kaji. Ilm. Inform. Dan Komput.*, vol. 3, no. 6, pp. 1340–1348, 2023.

[41] R. N. Ikhsani and F. F. Abdulloh, "Optimasi SVM dan Decision Tree Menggunakan SMOTE Untuk Mengklasifikasi Sentimen Masyarakat Mengenai Pinjaman Online," *J. Media Inform. Budidarma*, vol. 7, no. 4, pp. 1667–1677, 2023.

[42] W. Purbasari, N. Setianti, and O. Krismonika, "Analisis Sentimen dan Analisis Jaringan (Network Analysis) Seks Pranikah di Indonesia Menggunakan Data Media Sosial Twitter," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 12, Oct. 2023, doi: 10.30591/smartcomp.v12i4.5671.

[43] L. Legito *et al.*, "Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia: Implementation K-Nearest Neighbor Algorithm for Sentiment Analysis on Khilafah and Radicalism Issues in Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, pp. 324–330, Nov. 2023, doi: 10.57152/malcom.v3i2.893.

[44] M. F. Setiamukti and M. F. Nasvian, "Social Network Analysis #Usuttuntas Pada Media Sosial Twitter (Data Twitter 11 November 2022)," *Ekspresi Dan Persepsi J. Ilmu Komun.*, vol. 6, no. 1, pp. 124–137, 2023.

[45] Primandani Arsi, Pungkas Subarkah, and Bagus Adhi Kusuma, "Analisis Sentimen Game Genshin Impact pada Play Store Menggunakan Naïve Bayes Clasifier," *J. Ilm. Tek. Mesin Elektro Dan Komput.*, vol. 3, no. 1, pp. 161–170, Mar. 2023, doi: 10.51903/juritek.v3i1.1962.

[46] F. Rohmansyah and E. Poerwandono, "Analisis Data Sentimen Perbandingan Terhadap Game Online Mobile Legends dan PUBG Mobile Berdasarkan Tanggapan Masyarakat X Menggunakan Algoritma Naïve Bayes," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 7, pp. 1581–1587, Sep. 2024, doi: 10.31539/intecoms.v7i5.11824.

[47] S. Agung, "Implementasi Text Mining untuk Analisis Review pada Aplikasi Crowdfunding LX dan ST Menggunakan Metode Sentiment Analysis," *LANCAH J. Inov. Dan Tren*, vol. 2, no. 1, pp. 124–130, 2024.

[48] Novi Lestari, Elin Haerani, and Reski Candra, "Analisa Sentimen Ulasan Aplikasi WeTV Untuk Peningkatan Layanan Menggunakan Metode Naïve Bayes," *J. Inf. Syst. Res. JOSH*, vol. 4, no. 3, Apr. 2023, doi: 10.47065/josh.v4i3.3355.

[49] F. A. Indriyani, A. Fauzi, and S. Faisal, "Analisis Sentimen Aplikasi TikTok Menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *TEKNOSAINS J. Sains Teknol. Dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023.

[50] C. Zai and A. Isnain, "Komparasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) pada Analisis Sentimen Capcut," *INOVTEK Polbeng - Seri Inform.*, vol. 9, no. 1, 2024, doi: 10.35314/isi.v9i1.4054.

[51] A. Setyaningsih, D. Septiyani, and S. Widiasari, "Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia," *J. Teknol. Inform. Dan Komput.*, vol. 9, pp. 224–235, May 2023, doi: 10.37012/jtik.v9i1.1409.

[52] R. A. Permana and S. Sahara, "Algoritma K-Nearest Neighbor Pada Analisa Sentimen Review Produk Router," *J. Sist. Inf. Dan Sist. Komput.*, vol. 8, no. 2, pp. 118–124, 2023.

[53] R. A. E. Virgana Sapanji, Dani Hamdani, and Parlindungan Harahap, "Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia Using Text Mining and Natural Language Processing (NLP)," *J. Appl. Inform. Comput.*, vol. 7, no. 2, Nov. 2023, doi: 10.30871/jaic.v7i2.6517.

[54] A. Pramita and F. Nugraha, "Sistem Analisis Sentimen Produk Pada Aplikasi Lazada Menggunakan Metode Naïve Bayes," *J. Digit*, vol. 14, p. 23, Jun. 2024, doi: 10.51920/jd.v14i1.362.

[55] F. Wulandari, E. Haerani, M. Fikry, and E. Budianita, "Analisis Sentimen Larangan Penggunaan Obat Sirup Menggunakan Algoritma Naive Bayes Classifier," *J. CoSciTech Comput. Sci. Inf. Technol.*, vol. 4, no. 1, pp. 88–96, 2023.

[56] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes," *J. Sains Dan Teknol.*, vol. 5, no. 1, pp. 115–122, Aug. 2023.

[57] Y. Akbar, A. Regita, Sugiyono, and T. Wahyudi, "Analisa Sentimen Pada Media Sosial'X' Pencarian Keyword ChatGPT Menggunakan Algoritma K-Nearest Neighbors (KNN)," *J. Indones. Manaj. Inform. Dan Komun.*, vol. 5, pp. 3291–3305, Sep. 2024, doi: 10.35870/jimik.v5i3.1016.

[58] D. Atmajaya, A. Febrianti, and H. Darwis, "Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter," *Indones. J. Comput. Sci.*, vol. 12, Aug. 2023, doi: 10.33022/ijcs.v12i4.3341.

[59] A. A. Munandar, F. Farikhin, and C. E. Widodo, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM," *JOINTECS J. Inf Technol Comput Sci Vol 8 No 2 P 77 2023 Doi 1031328jointecs V8i2 4747*, 2023.

[60] R. Husen, R. Astuti, L. Marlia, R. Rahmaddeni, and L. Efrizoni, "Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning: Sentiment Analysis of Public Opinion on Twitter Toward BSI Bank Using Machine Learning Algorithms," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, pp. 211–218, Oct. 2023, doi: 10.57152/malcom.v3i2.901.

[61] Kurnia, I. Purnamasari, and D. Saputra, "Analisis Sentimen Dengan Metode Naïve Bayes, SMOTE Dan Adaboost Pada Twitter Bank BTN," *J. JTIK J. Teknol. Inf. Dan Komun.*, vol. 7, pp. 235–242, Apr. 2023, doi: 10.35870/jtik.v7i3.707.

[62] S. Saropah, R. Astuti, and F. M. Basysyar, "Algoritma Naïve Bayes Untuk Melakukan Analisa Sentimen Terhadap Aplikasi Axisnet Di Google Play," *JATI J. Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3655–3660, 2023.

[63] S. Butsianto, S. Fauziah, C. Naya, and F. Maulana, "Sentiment Analysis Of Indosat's Mobile Operator

Services On Twitter Using The Naïve Bayes Algorithm," *Brill. Res. Artif. Intell.*, vol. 4, no. 1, pp. 245–254, Jun. 2024, doi: 10.47709/brilliance.v4i1.4084.

[64] F. Syah, H. Fajrin, A. Afif, M. Saeputra, D. Mirranty, and D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *J. JTIK J. Teknol. Inf. Dan Komun.*, vol. 7, pp. 53–58, Jan. 2023, doi: 10.35870/jtik.v7i1.686.

[65] D. Pratmanto, R. Rousyati, and A. Widodo, "Analisa Sentimen Persepsi Masyarakat Terhadap Aplikasi Bea Cukai Mobile Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors," *EVOLUSI J. Sains Dan Manaj.*, vol. 12, Sep. 2024, doi: 10.31294/evolusi.v12i2.23576.

[66] Ratih Puspitasari, Y. Findawati, and M. A. Rosid, "Sentiment Analysis Of Post-Covid-19 Inflation Based On Twitter Using The K-Nearest Neighbor And Support Vector Machine Classification Methods," *J. Tek. Inform. Jutif*, vol. 4, no. 4, pp. 669–679, Aug. 2023, doi: 10.52436/1.jutif.2023.4.4.801.

[67] A. D. Cahyani, "Analisa Kinerja Metode Support Vector Machine untuk Analisa Sentimen Ulasan Pengguna Google Maps," *J. Comput. Syst. Inform. JoSYC*, vol. 4, no. 3, pp. 604–613, 2023.

[68] E. Tohidi, R. P. Herdiansyah, E. Wahyudin, and K. Kaslani, "Analisa Sentimen Komentar Video Youtube Di Channel Tvonenews Tentang Calon Presiden Prabowo Subianto Menggunakan Support Vector Machine," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 660–667, 2024.

[69] C.-H. Lin and U. Nuha, "Sentiment Analysis of Indonesian Datasets Based on A Hybrid Deep-Learning Strategy," *J. Big Data*, vol. 10, no. 1, p. 88, May 2023, doi: 10.1186/s40537-023-00782-9.

[70] R. Sukmana and A. Rusydiana, "Social Media Sentiment Analysis on Waqf and Education," *Islam. Mark. Rev.*, vol. 2, Dec. 2023, doi: 10.58968/imr.v2i2.325.

[71] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, 2017, pp. 391–394. doi: 10.1109/IALP.2017.8300625.

[72] H. Imaduddin, F. A'la, and Y. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, Jan. 2023, doi: 10.14569/IJACSA.2023.0140813.

[73] N. Royanti, I. Indrayanti, and B. Ismanto, "Analisa Sentimen Pelanggan pada Review Belanja Online Berbasis Text Mining Menggunakan Metode K-Means," *J. Inf. Syst. Res. JOSH*, vol. 4, pp. 1441–1447, Jul. 2023, doi: 10.47065/josh.v4i4.3781.

[74] A. Adam and E. Setiawan, "Social Media Sentiment Analysis using Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU)," pp. 119–131, Mar. 2023, doi: 10.26555/jiteki.v9i1.25813.

[75] H. Setyawan, L. Azizah, and A. Pradani, "Sentiment Analysis of Public Responses on Indonesia Government Using Naïve Bayes and Support Vector Machine," *Emerg. Inf. Sci. Technol.*, vol. 4, pp. 1–7, May 2023, doi: 10.18196/eist.v4i1.18681.

[76] Y. Findawati, U. Indahyanti, Y. Rahmawati, and R. Puspitasari, "Sentiment Analysis of Potential Presidential Candidates 2024: A Twitter-Based Study," *Acad. Open*, vol. 8, Aug. 2023, doi: 10.21070/acopen.8.2023.7138.

[77] D. Era, S. Andryana, and A. Rubhasy, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19," *J-SAKTI J. Sains Komput. Dan Inform.*, vol. 7, no. 1, pp. 263–272, 2023.

[78] A. Halim and Andri Safuwan, "Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization," *J. Inform. Teknol. Dan Sains Jinteks*, vol. 5, no. 1, pp. 170–178, Feb. 2023, doi: 10.51401/jinteks.v5i1.2229.

[79] N. Habibah, E. Budianita, M. Fikry, and I. Iskandar, "Analisis Sentimen Mengenai Penggunaan E-Wallet Pada Google Play Menggunakan Lexicon Based dan K-Nearest Neighbor," *JURIKOM J. Ris. Komput.*, vol. 10, no. 1, pp. 192–200, 2023.

[80] E. Nurraharjo and others, "Analisis Sentimen Dan Klasifikasi Tweet Terkait Naiknya Kasus Omicron Menggunakan Naive Bayes Classifier," *J. Inform. Dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 1–8, 2023.

[81] A. Fatkhudin, F. A. Artanto, N. A. Safli, and D. Wibowo, "Decision Tree Berbasis SMOTE dalam Analisis Sentimen Penggunaan Artificial Intelligence untuk Skripsi," *REMIK Ris. Dan E-J. Manaj. Inform. Komput.*, vol. 8, no. 2, pp. 494–505, 2024.

[82] M. R. Qisthiano, I. Ruswita, and P. A. Prayesy, "Implementasi Metode SVM dalam Analisis Sentimen Mengenai Vaksin dengan Menggunakan Python 3," *Teknol. J. Ilm. Sist. Inf.*, vol. 13, no. 1, pp. 1–7, 2023.

[83] M. R. Prasetyo and A. Fahrurozi, "Analisa Sentimen Pada Ulasan Google Untuk Hotel Gran Mahakam Jakarta Menggunakan Pendekatan Machine Learning," *J. Ilm. Inform. Komput.*, vol. 28, no. 3, pp. 203–217, 2023.

[84] Adetya Rizal Permana Putra Rizal and Jati Sasongko Wibowo, "Sentiment Analysis Twitter Sentiment Analysis of the 2024 Indonesian Presidential Candidates Using the KNN Method," *Elkom J. Elektron. Dan Komput.*, vol. 17, no. 1, Jul. 2024, doi: 10.51903/elkom.v17i1.1603.

[85] I Komang Andi Sugiarta, P. Anugrah Cahya Dewi, and Nengah Widya Utami, "Analisa Sentimen Mahasiswa Terhadap Layanan Stmik Primakara Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor," *J. Inform. Teknol. Dan Sains Jinteks*, vol. 5, no. 3, pp. 364–372, Aug. 2023, doi: 10.51401/jinteks.v5i3.3159.