Design of Sentiment Analysis on Indodax Instagram Social Media Comments About Cryptocurrency Using Naïve Bayes Classifier

Dimas Prayoga
Department of Information System, Universitas Muhammadiyah Sumatera Utara, Indonesia

ABSTRACT

In today's digital era, social media such as Instagram has become the main platform for many individuals to interact and express opinions online. One application that is often the subject of conversation is Indodax, a well-known digital asset trading platform in Indonesia. This research aims to evaluate the sentiment of Instagram users towards Indodax services through a sentiment analysis approach using Naive Bayes Classifier. The data collected consists of Instagram users' comments, which are analyzed to assess the tendency of their sentiments, whether positive or negative towards Indodax services. This method applies probability and statistical concepts to classify sentiments based on the words present in the comments. It is hoped that the results of this study can provide insights for Indodax to improve the quality of their services based on the perceptions of users. Based on the experiments conducted, the Naive Bayes Classifier method shows fairly accurate classification results, so it can support sentiment analysis related to the Indodax application.

Keyword : User Sentiment, Indodax, Instagram, Naive Bayes Classifier, Application Service Quality

Corresponding Author:
Dimas Prayoga,
Department of Information System,
Universitas Muhammadiyah Sumatera Utara,
Jalan Kapten Muktar Basri No 3 Medan 20238, Indonesia.
Email: dimasprayoga@gmail.com

1. INTRODUCTION

Social media has become a crucial element in the lifestyle of internet users in today's digital era. The function of social media as a means of interaction between internet users also acts as a communication tool in the context of friendship, family, and other relationships. Various interactions can be done through social media, such as voice calls, video calls, and text message exchanges (Efraim, 2023). Instagram is one of the most popular social media platforms for sharing ideas and interacting with others. For many individuals, Instagram is more than just a social network; it has become an important part of their daily lives, serving as a means for social interaction and creative expression (1978, Informatics).

Indodax (PT Indodax Nasional Indonesia) is a technology-based company that connects the largest digital asset sellers and buyers in Indonesia. Since operating in 2014, Indodax has served more than 4 million members in 80 countries and offers more than 160 tradable crypto assets. With over 10 million visitors and a trading volume of IDR 3 trillion per month, Indodax is known as a highly liquid platform, ranking fourth in the world in the crypto asset market for web traffic according to ICO Analytics in 2018 (Wikipedia).

Behind the use of the Indodax platform as the main provider for digital asset trading, there are various positive and negative assessments from users regarding the quality of services provided by the application. Many Indodax application users express comments about the quality of service on Instagram. These comments are used to conduct sentiment analysis to determine user tendencies, whether these comments tend to be positive or negative towards Indodax services.

Public opinion sentiment analysis is the process of understanding people's views, feelings, and attitudes towards a particular cryptocurrency trading application or service. In this context, researchers will use comment data from Instagram users regarding their experiences in using Indodax services. However, there are several challenges that must be faced, such as the use of abbreviations and slang in comments. Therefore, researchers will conduct sentiment analysis regarding public responses. By

utilizing Instagram comment data, we can evaluate whether the application is worthy as a financial service platform (Industry & Indonesia, 2023). This study aims to measure Instagram user sentiment towards Indodax services in Indonesia (Zendrato et al., 2024). Through sentiment analysis techniques, researchers will identify whether the comments tend to be positive or negative, providing valuable insights for Indodax to understand user perceptions and improve their experiences in the future (Zendrato et al., 2024).

This study aims to offer a solution to customer sentiment analysis of Indodax services using the Naive Bayes Classifier method. This method is used to classify customer sentiment based on comments given on Instagram. Naive Bayes Classifier utilizes probability and statistics to group certain variables into positive or negative sentiment categories, including the application of variants such as Complement Naive Bayes, Multinomial Naive Bayes, and Bernoulli Naive Bayes. This method is not only used to classify customer sentiment but also to produce test results with a high level of accuracy. Naive Bayes was developed by Reverend Thomas Bayes in the 18th century, and classification with the Naive Bayes method is generally carried out through a chance or probability approach (Sipayung et al., 2016).

2. RESEARCH METHOD

This research is an experimental study that uses data collected independently. This scientific work focuses on the analysis of moments with the aim of identifying positive and negative moments. The initial step in this research is to collect data from Instagram social media users through keyword searches on moments that are relevant to the topic being studied. The process of collecting data in the form of moments is also known as data crawling. In this study, the software for text analysis used is Python, equipped with libraries such as NLTK (Natural Language Toolkit) to perform data pre-processing and implement the Naive Bayes Classifier method. The data used in this study comes from intelligence on cryptocurrency posts published by the Indoidax official account on Instagram. Data collection was carried out using OipeinReifinei and IGCoimmeintEixpoirt. In this dataset, the attributes contained include intelligence text and username. The collection process begins by accessing posts related to cryptocurrency on the Indoidax Instagram account. Furthermore, the keywords in the post are extracted using OipeinReifinei and IGCoimmeintEixpoirt, then stored in the form of a dataset. Each keyword is accompanied by the user attribute user name that contains the keyword and the contents of the keyword text itself. After the data collection stage is complete, the next step is to conduct an analysis of the keywords using the Naive Bayes Classifier. This method was chosen because it is effective in classifying keywords based on text. The Naive Bayesian model will be trained using a dataset of intelligent coins that have been labeled with integers (positive and negative). After training, the model will be used to classify integers from new, unlabeled intelligent coins. Through the analysis of integers on intelligent coins on Instagram Indoidax related to cryptocurrency, it is hoped that valuable insights can be obtained regarding user views and opinions on cryptocurrency as well as an understanding of the trends and patterns of integers that can influence the market.

3. RESULTS AND DISCUSSION

A. Data Presentation

Data presentation is a crucial step in analysis, with the goal of conveying information clearly through effective visualization. This process involves selecting appropriate visualization methods, such as graphs, charts, or maps, to depict data in a clear and informative manner. Data presentation makes it easier to interpret the results of the analysis, allowing users to understand patterns, trends, and insights contained in the data. Using the right visualization tools in data presentation helps simplify complex data, thereby speeding up and simplifying the decision-making process.

1. Data Selection

The data selection stage is a key phase in the analysis that focuses on selecting relevant data from the collection of comments on Indodax Instagram posts. In this stage, the collected data is filtered and selected using certain criteria to ensure that only the necessary data is used in subsequent analysis. This process involves selecting relevant variables, such as comment date, comment text, and sentiment, and removing incomplete, duplicate, or inappropriate data. In addition, data selection also includes determining a representative subset of data from a larger population for more efficient analysis. With proper data selection, we can ensure that subsequent analysis is more focused and effective, resulting in sharper insights and better data-based decision making.

Fig 1. Data Selection

2. Data Scraping

The data scraping process to collect comments from the Instagram platform can be done by utilizing Instagram post comments automatically. The purpose of this project is to analyze comments related to Indodax services about cryptocurrency, such as products or marketing campaigns. Data scraping The data selection and visualization stages are carried out using the Python programming language, with the help of several popular libraries. The Pandas library is used for data manipulation, facilitating filtering, removing irrelevant data, and selecting variables. For visualization purposes, Matplotlib and Plotly are used to present data in the form of informative graphs. In natural language processing (NLP), NLTK is used for tokenization, stemming, and removing stopwords. In addition, WordCloud helps display visualizations of the most frequently occurring words in the form of word clouds, so that dominant keywords can be easily identified in comment data. Jupyter Notebook and Google Colaboratory are often used as platforms for running Python scripts, as they provide an integrated environment and support data analysis.

The data collected came from Instagram comments during the period of May, from 07-05-2024 to 01-08-2024. Scraping was done using libraries such as BeautifulSoup or Selenium, which allow for automatic data collection from websites by extracting HTML content. This process involves taking data directly from the Indodax Instagram page, focusing on relevant comment text. After the data was collected, preprocessing was carried out by removing unnecessary columns and focusing on the comment text. This process involved cleaning the text using regex to remove unwanted characters and tokenization to break the text into individual words. Stop words in Indonesian were removed to improve analysis efficiency, and stemming and lemmatization techniques were applied using NLTK to simplify word forms.

This data scraping provides advantages over manual methods, such as speed and efficiency in data collection. However, this data collection process also requires a deep understanding of the structure of web pages and techniques for handling dynamic content on Instagram. Automation techniques, such as web scraping programming or the use of APIs, help access and extract large amounts of data quickly. This approach allows authors to collect data with high accuracy and relevance, speeding up the overall

analysis process, and increasing efficiency in identifying patterns or trends that emerge in user comments on Instagram.

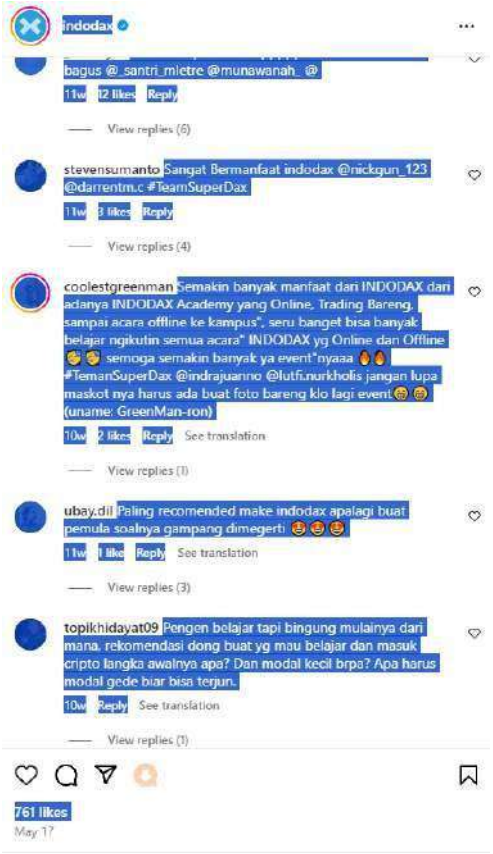

Fig 2. Comment Scraping

Fig 3. Comment Scraping Result

3. Open Refine

The data processing stage using OpenRefine is a very important step, which involves a number of processes to clean and tidy up the data before it is analyzed or used in further modeling. This process begins with importing data from various sources, such as CSV or Excel files. After that, data cleaning is carried out, including removing duplicate entries, fixing inconsistencies in format, and handling missing values. The clustering feature is used to identify and unite similar but differently written values. After that, data transformation is carried out to modify or rearrange the data according to the analysis needs.

This process ends with the export of the cleaned data to the desired format. All of these stages can be done with the intuitive, web-based OpenRefine interface. Data processing with OpenRefine is very important because it allows the handling of complex and unstructured data to be more organized and ready for analysis.

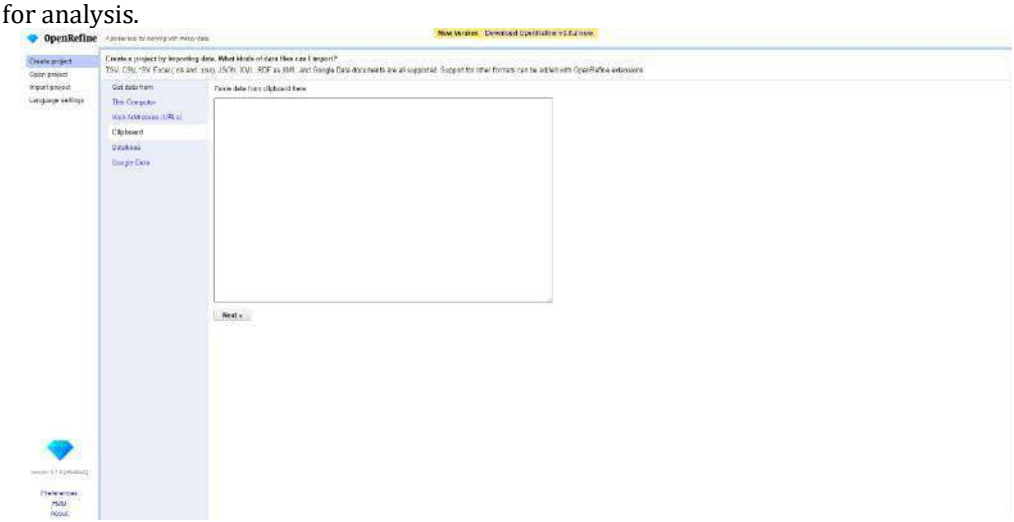

Fig 4. Open Refine Dashboard

B. Data Processing

The data processing stage is a crucial process that includes several steps to prepare and tidy up the data before it is used in further analysis or modeling. In text data processing, several important steps start with case folding, which converts all letters to lowercase for consistency. Next, the cleaning process is carried out to remove irrelevant characters, including punctuation and numbers. After that, tokenization breaks the text into small units called tokens. The next step is stopword removal, which eliminates common words that do not provide analytical value, such as "and", "or", and "which". Then, normalization is carried out to standardize the text, ending with stemming, which converts words to their basic form. Google Collaboratory and the Python programming language can be used for all of these tasks. Because raw data is often unstructured and unstable, it requires organization so that algorithmic models can understand it effectively, making this data processing step important.

1. Case Folding

The first step in the text data processing process in this study is case folding. This process aims to change all characters in the dataset to lowercase. By changing all text to lowercase, we can ensure that the data is more consistent and easier to process, because it avoids differences in interpretation that may arise due to the use of uppercase and lowercase letters in the text. This process is an important step in the broader text processing stages, including cleaning, tokenization, stopword removal, normalization, and stemming, all of which aim to improve the quality and consistency of the data before further analysis.

After performing case folding, the results of the text data in the dataset have been converted to lowercase as follows:

Table 1. Case Folding Result

Full Text	Case Folding Result
@sukasis_ INDODAX exchange yang terpercaya dan simpel te…	@sukasis_ indodax exchange yang terpercaya dan simpel te...
@lyptra_ Awal saya kenal indodax ini karena Timothy ron...	@lyptra_ awal saya kenal indodax ini karena timothy ron...
@coolestgreenman Semakin banyak manfaat dari INDODAX dari adany...	@coolestgreenman semakin banyak manfaat dari indodax dari adany...

2. Tokenizing

Tokenization is the next phase that involves dividing the sentence into separate words, or tokens, to make it easier to identify the root words of the phrase.

Regex from the Python re package is used for tokenization, not nltk. The re.split function creates a list of tokens (words) in each comment by splitting the text by non-alphabetic characters (\W+). The data taken for the tokenization process comes from the Comments column, which contains the text of the comments. The result of this tokenization is a list of words contained in each sentence in the Comments column. Furthermore, the tokenization results are stored in a new column called Tokenized_Comments. This code also displays the first 30 rows of the Tokenized_Comments column to see the tokenization results.

Table 2. Tokenizing Result

Before separating the text	After separating the text
indodax my exchange terbaik menurut gua ngga a...	[indodax, my, exchange, terbaik, menurut, gua,...
awal saya kenal indodax ini karena timothy ron...	[awal, saya, kenal, indodax, ini, karena, timo...
alhamdulillah selama pakai indodax sangat nyam...	[alhamdulillah, selama, pakai, indodax, sangat...
pengen belajar tapi bingung mulainya dari mana...	[pengen, belajar, tapi, bingung, mulainya, dar...

3. Normalization

Normalization is an important step in natural language processing (NLP) that aims to simplify and unify the various forms of words in text into a consistent standard form. This process involves several key steps, such as lowercasing, which is changing all characters in the text to lowercase to avoid the difference between uppercase and lowercase letters. In addition, stemming is applied to simplify words to their basic form by removing suffixes or prefixes, for example changing "running" and "runner" to "run".

	A	B	C
1	aktip	aktif	
2	aktifitas	aktivitas	
3	Apotik	Apotek	
4	apotik	Apotek	
5	analisa	analisis	
6	azas	asas	
7	Azas	asas	
8	atlit	Atlet	
9	Atlit	Atlet	
10	Antri	antre	
11	antri	antre	
12	Atmosfir	Atmosfer	
13	atmosfir	Atmosfer	
14	Erobik	Aerobic	
15	erobik	Aerobic	

Fig 8. Normalization Dictionary

Lemmatization is a more advanced step, which not only changes the word to its base form but also considers its meaning and context, for example, changing "better" to "good". The process of removing punctuation is also important to remove unnecessary punctuation from the text. By applying normalization, we can ensure that the text is processed in a consistent form and facilitate further analysis, such as data processing and sentiment analysis.

Table 3. Normalization Result

Normalization

indodax keren jaya selalu makin rame usernya t...
aplikasi jual beli kripto yang sangat mudah di...
aplikasi indodax ini sangat bermanfaat untuk y...

4. Filtering – Removing Stopwords

The next step is the removal of stopwords, which is the process of filtering common words that do not have important meanings in the text. These words often appear in large numbers but do not contribute significantly to the core understanding of the content. Examples of such words in English include "the," "is," "at," "which," and so on.

This process is useful for simplifying the text by eliminating unnecessary words, so that the analysis can be more focused on words that have meaning. Thus, the process of identifying the class or category of the text will be easier and more accurate.

Stopwords Removal - Filtering can be done by utilizing libraries such as NLTK (Natural Language Toolkit) in the Python programming language, which provides a list of standard stopwords that can be used as a reference for the filtering process. In the context of text analysis or natural language processing, stopword removal helps improve the efficiency and accuracy of the model by reducing noise in the data.

Table 4. Filtering Result

Filtering
indodax keren favorit aplikasi aplikasi nya me...
indodax keren jaya rame usernya teamsuperdax m...
1 bulanan buka akun akomodatif terima kasih in...

5. Stemming

The next stage is Stemming, a technique in text preprocessing that functions to reduce words to their basic form or root words. This process is crucial in natural language processing because it allows grouping variations of word forms that have similar meanings. For example, words like "running," "running-running," and "runner" will be simplified to the same basic form, namely "run." By removing unnecessary endings or prefixes, stemming allows the system to process text more efficiently and reduces the complexity of the analysis.

In practice, stemming uses a specific algorithm, such as the Porter Stemmer or Snowball Stemmer, to identify and remove affixes from words. The result of stemming is not a valid word in everyday language, but rather a general representation of the word that is sufficient for the purposes of text analysis. Although stemming can improve the efficiency of search and analysis, it should be noted that it can sometimes produce inaccurate or confusing word forms, especially in languages with many morphological rules.

Table 5. Data Stemming Result

Stemming Result
jual beli bitcoin aset kripto indodax segampan...
guru sekolah main buka fitur trade beliau mene...
indodax terbaik pelayanan cepat ramah temansup...

C. Sentiment Analyze

Sentiment analysis in this code is done in a simple way, namely by counting the number of positive and negative keywords that appear in comments. The analyze_sentiment function calculates the frequency of occurrence of keywords from the positive_keywords and negative_keywords lists in comments. Based on this calculation, the sentiment of the comment is categorized as 'positive' if there are more positive keywords, 'negative' if there are more negative keywords, or None for neutral sentiment if the number of positive and negative keywords is the same. This code then adds a 'Sentiment' column to the DataFrame and removes rows with neutral sentiment to clarify the analysis.

This approach mirrors lexicon-based methods in sentiment analysis, where sentiment is determined by the frequency of pre-categorized keywords. While this method is efficient in providing an initial picture of sentiment, it has limitations in capturing more complex context or emotional nuances in comments, such as sarcasm or double meanings. Therefore, the results of this analysis may need to be combined with other techniques to gain a deeper and more accurate understanding of sentiment.

Table 8. Sentiment Analyze Result

User	Comment	Sentiment
Rupifebriani	kak emang indodax tf	positif
candr_a2115	1juta pencairan 2juta	positif
fajarimantoro_	ver... 1thun main indodax bermanfaat... indodax gabisa kebuka ya	negatif

D. Naïve Bayes Modelling

Confusion Matrix is a crucial evaluation tool in machine learning used to measure the performance of a classification model. This matrix displays the comparison between the model predictions and the actual labels in a tabular form. By including information about true positives, false positives, true negatives, and false negatives, the Confusion Matrix provides a clear picture of the model's errors and successes in classifying data into the correct class. Analysis of this matrix makes it easy to understand where the model might be making mistakes and how it performs in each class. This information is adapted with modifications from sources available on Kaggle, specifically from the project on sentiment analysis using Naive Bayes Classifier by Ankumagawa (Kaggle, 2024).

Table 9. Confusion Matrix

Actual/Prediction	Positive Prediction	Negative Prediction
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Explanation:

- True Positive (TP): Refers to the number of cases that are actually positive and also predicted positive by the model.
- False Negative (FN): The number of cases that are actually positive but predicted negative by the model.
- False Positive (FP): Indicates the number of cases that are actually negative but predicted positive by the model.
- True Negative (TN): Indicates the number of cases that are actually negative and also predicted negative by the model.

E. Analysis Result

In this study, data was collected through the web scraping method from the social media source Instagram indodax. The data collection process was carried out using the Open Refine, Jupyter Notebook, and Google Colaboratory tools with the Python programming language. The collected data was then further processed to ensure its accuracy and completeness before entering the analysis stage.

The collected data then went through a pre-processing process to clean it and prepare it for use in further analysis. The pre-processing stages include steps such as case folding, text cleaning,

tokenization, normalization, stopword removal, and stemming. After this process, the data becomes more structured and ready to be used in classification.

The processed data is then labeled based on the sentiment contained. In this study, three sentiment classes are applied, namely positive, negative, and neutral. The labeling process is carried out using a lexicon approach, where the data is compared to a previously compiled lexicon dictionary. The result of this process is data that is classified according to the sentiment that has been identified.

Three classification algorithms are used in this study to categorize the data: Multinomial Naive Bayes, Complement Naive Bayes, and Bernoulli Naive Bayes. The selection of these three algorithms is based on their respective advantages in classifying text data with diverse characteristics, thus allowing for comparative evaluation of performance between methods.

After the classification is complete, an evaluation is carried out to assess the performance of each algorithm. This evaluation process includes measurements using a confusion matrix as well as calculating accuracy, precision, recall, and F1-score values. The results of the evaluation provide insight into how well the models perform the correct classification.

From the evaluation results, the Multinomial Naive Bayes model managed to record an accuracy of 81%, followed by Complement Naive Bayes with an accuracy of 77%, and Bernoulli Naive Bayes which obtained an accuracy of 76%. These results indicate that Complement Naive Bayes has a better performance in data classification in this study, while Bernoulli Naive Bayes has a slightly lower performance than the other two models.

In addition to measuring accuracy, other metrics such as precision, recall, and F1-score are also calculated for each sentiment class. The results of the analysis show that Complement Naive Bayes shows consistent and superior performance in various metrics, followed by Bernoulli Naive Bayes after Multinomial Naive Bayes. Despite its lower accuracy, Bernoulli Naive Bayes still shows good performance in certain situations, especially when faced with data with binary features.

Based on the evaluation results, the researcher concluded that Complement Naive Bayes is the most effective algorithm for sentiment classification in the context of this study, followed by Multinomial Naive Bayes, and finally Bernoulli Naive Bayes. The results of the sentiment analysis also show that most of the data tends to be negative sentiment, indicating a negative perception from users towards the topic being analyzed.

4. CONCLUSION

This study aims to analyze Instagram user sentiment towards Indodax services in Indonesia through a text analysis approach. The process begins with collecting relevant comment data using web scraping techniques, then continues with a data cleaning stage that includes normalization, tokenization, and stemming to ensure data quality before analysis is carried out. After the data is prepared, the comments are classified into two sentiment categories, namely positive and negative, using a lexicon approach. For classification analysis, three variants of the Naive Bayes algorithm are used, namely Multinomial Naive Bayes, Complement Naive Bayes, and Bernoulli Naive Bayes, each of which has advantages in handling word distribution and data imbalance. The test results show that Multinomial Naive Bayes is able to provide the best performance with an accuracy of 81%, followed by Complement Naive Bayes with an accuracy of 77%, and Bernoulli Naive Bayes at 76%. These findings indicate that Multinomial Naive Bayes is more effective in managing text data and providing more accurate sentiment classification results, so it is recommended as the best method for analyzing user comment sentiment towards Indodax services on social media, especially Instagram.

REFERENCES

[1] Alifta, R. F. (2022). UNIKOM_RadiffaFizryAlifta_BAB II (pp. 9–30).

[2] Anadakumar, K., & Padmavathy, V. (2013). Tinjauan tentang Proses Preprocessing dalam Text Mining. *International Journal of Advanced Research in Computer Science*, 4(9), 79–91.

[3] Anwar, K. (2022). Analisis Sentimen Pengguna Instagram di Indonesia terkait Ulasan Smartphone dengan Menggunakan Naive Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 2(4), 148–155. https://doi.org/10.30865/klik.v2i4.315

[4] Asyrofi, R. R., & Asyrofi, R. (2023 Penggunaan Aplikasi Jupyter Notebook dalam Menganalisis Kriteria Plagiasi dengan Metode Semantik. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(2), 627–637. https://doi.org/10.29100/jipi.v8i2.3699

[5] Bagus, M., Sukoco, A., Informasi, T., Informatika, T. D., Bina, U., Informatika, S., Selatan, K. T., & Chaum, D. (2024). Dampak Metode PIECES pada Transaksi dalam Aplikasi. *Jurnal Ilmiah*, 8(3), 3339–3342.
[6] Berniawan, D., Amri, A., & Tinaliah, T. (2023). Penerapan Algoritma Naïve Bayes untuk Klasifikasi Sentimen Pengguna Twitter terhadap KEMKOMINFO di Indonesia. *MDP Student Conference*, 2(1), 24–31. https://doi.org/10.35957/mdp-sc.v2i1.4326
[7] Chuzaimah Zulkifli, U. (2018). Pengembangan Modul Preprocessing Teks untuk Memformalkan dan Memeriksa Ejaan Bahasa Indonesia dalam Aplikasi Web Mining Simple Solution (WMSS). *Jurnal Matematika Statistika dan Komputasi*, 15(2), 95. https://doi.org/10.20956/jmsk.v15i2.5718
[8] Daryfayi, E., Daulay, P., & Asror, I. (2020). Analisis Sentimen pada Ulasan Google Play Store dengan Metode Naïve Bayes. *E-Proceeding of Engineering*, 7(2), 8400–8410.
[9] Dianti, Y. (2017). Judul Tidak Diketahui. *Angewandte Chemie International Edition*, 6(11), 951–952. http://repo.iain-tulungagung.ac.id/5510/5/BAB 2.pdf
[10] E-servqual, M. M. (2021). Analisis Kualitas Layanan Aplikasi Indodax dengan Metode E-Servqual dan Importance Performance Analysis (IPA). *Jurnal Ilmiah Komputasi*, 20(3), 425–433. https://doi.org/10.32409/jikstik.20.3.2735
[11] Efraim, D. A. (2023). Analisis Sentimen pada Media Sosial Instagram Menggunakan Algoritma Naive Bayes (Studi Kasus: Timnas Futsal Indonesia). *Jurnal Ilmiah*, 498–509.
[12] Eugen, E. (2021 Visualisasi Web Mining untuk Popularitas Sembilan Universitas Swasta Terbaik di Jakarta. 74. https://kc.umn.ac.id/15921/
[13] Fahlevvi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine. *Jurnal Teknologi dan Komunikasi Pemerintahan*, 4(1), 1–13. https://doi.org/10.33701/jtkp.v4i1.2701
[14] Ghosh, S., Hazra, A., & Raj, A. (2020). Studi Perbandingan Berbagai Teknik Klasifikasi dalam Analisis Sentimen. *International Journal of Synthetic Emotions*, 11(1), 49–57. https://doi.org/10.4018/ijse.20200101.oa
[15] Informatika, J. T. (1978). Implementasi Naive Bayes Classifier dalam Menganalisis Sentimen Komentar Pelanggan Mie Gacoan di Instagram. *Jurnal Ilmiah*, 18(x), 139–149.
[16] Iskandar, I., Muhadar, M., & Hijrah, H. (2021). Peran Unit Investigasi Kriminal di Polda Sulawesi Selatan dalam Penegakan Hukum Terhadap Penyebar Hoaks. *Jurnal Cakrawala Hukum*, 12(3), 284–293. https://doi.org/10.26905/idjch.v12i3.5159
[17] Karim, A. (2020). Analisis Sentimen pada Komentar di Media Sosial Instagram Layanan Kesehatan BPJS Menggunakan Naive Bayes Classifier. *Skripsi*, 5(3), 248–253.
[18] Khatibah, K. (2011). Jurnal Perpustakaan dan Informasi. *Iqra'*, 2275(Penelitian Kepustakaan), 36–39.
[19] Mahardika, Y. S., & Zuliarso, E. (2018). Analisis Sentimen terhadap Pemerintahan Joko Widodo di Media Sosial Twitter dengan Menggunakan Algoritma Naïve Bayes Classifier. *Prosiding SINTAK 2018*, 2015, 409–413.
[20] Mathapati, P. M., Shahapurkar, A. S., & Hanabaratti, K. D. (2017). Analisis Sentimen dengan Menggunakan Algoritma Naïve Bayes. *International Journal of Computer Sciences and Engineering*, 5(7), 75–77. https://doi.org/10.26438/ijcse/v5i7.7577
[21] Muhammad Romzi, & Kurniawan, B. (2020). Pembelajaran Pemrograman Python melalui Pendekatan Logika Algoritma. *JTIM: Jurnal Teknik Informatika Mahakarya*, 03(2), 37–44.
[22] Nuzulia, A. (1967). 済無 No Title No Title No Title. *Angewandte Chemie International Edition, 6(11), 951–952.*
[23] Patappari, A., & Syafei. (2021). Perancangan Aplikasi Penyewaan Ruang Pertemuan. Jisti, 4, 39–49.
[24] Pradana, M. G. (2020). Penggunaan Fitur Wordcloud dan Document Term Matrix dalam Text Mining. Jurnal Ilmiah Informatika, 8(1), 38–43.
[25] Pustaka, T. (2023). *Tahun Publikasi. Jurnal Ilmiah, 7(1), 681–686.*
[26] Rifaldi, D., Abdul Fadlil, & Herman. (2023). Teknik Preprocessing dalam Text Mining dengan Menggunakan Data Tweet "Mental Health." Decode: Jurnal Pendidikan Teknologi Informasi, 3(2), 161–171. https://doi.org/10.51454/decode.v3i2.131
[27] Rizkya, A. T., Rianto, R., & Gufroni, A. I. (2023). Penerapan Naive Bayes Classifier untuk Analisis Sentimen pada Data Ulasan Aplikasi E-Commerce Shopee di Google Play Store. International Journal of Applied Information Systems and Informatics (JAISI), 1(1), 31–37.
[28] *Scaling OpenRefine*. (2019). Halaman 1–9.

[29] Septiani, D., & Isabela, I. (2022). Analisis Term Frequency Inverse Document Frequency (Tf-Idf) dalam Temu Kembali Informasi pada Dokumen Teks. SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia, 1(2), 81–88.

[30] Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan dengan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi (JSI)*, 8(1), 2355–4614. http://ejournal.unsri.ac.id/index.php/jsi/index

[31] Taufik, M., Harahap, E. N., & Akbar, H. A. (2020). Analisis Sentimen pada Ulasan Aplikasi Berbasis Android dengan Algoritma Naive Bayes. Jurnal Analisis Data dan Teknologi Informasi, 7(1), 15–22.

[32] Wibawa, I., & Dewi, A. (2021). Analisis Sentimen pada Ulasan Game Mobile Legends dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Computer Science & Information Technology*, 6(2), 112–119. https://doi.org/10.31334/jcst.v6i2.928

[33] Wibowo, R., & Abdul Aziz. (2023). Analisis Sentimen di Instagram untuk Klasifikasi Ulasan Makanan Menggunakan Metode Naïve Bayes Classifier. *Jurnal Komputer*, 21(1), 135–141. https://doi.org/10.30865/jk.v21i1.2907

[34] Tandel, S. S., Jamadar, A., & Dudugu, S. (2019 Survei Teknik Text Mining. *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 1022–1026. https://doi.org/10.1109/ICACCS.2019.8728431

[35] Zendrato, A. D., Berutu, S. S., Sumihar, P., & Budiati, H. (2024). *Pengembangan Model Klasifikasi Sentimen Dengan Pendekatan Vader dan Algoritma Naive Bayes Terhadap Ulasan Aplikasi Indodax*. *5*(3). https://doi.org/10.47065/josh.v5i3.5050

[36] Zain, I., Jamil, A., & Zahratul D. (2022). Implementasi Metode Naive Bayes untuk Klasifikasi Data Ulasan. *Jurnal UMSurabaya*, 10(1), 41–50. https://doi.org/10.30736/um-skripsi.v10i1.2616