Research Article

Sentiment analysis of Faculty of Science and Technology students' satisfaction with the 2024 graduation using the Naïve Bayes method

[1,*]Kalfida Eka Wati Siregar, [2]Wily Supi Ramadani, [3]Anggi Jelita Sitepu, [4]Ulfi Muzayyanah Fadil, [5]Mhd. Furqan

[1,2,3,4,5] Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Medan, North Sumatera, Indonesia

* Corresponding Author: kalfidaekawatisiregar@gmail.com

Abstract: Sentiment analysis of UINSU student graduation based on academic data is one of the efforts to understand the factors that affect the success of student studies. This research aims to analyze the sentiment of UINSU student graduation by utilizing academic data such as cumulative grade point average (GPA), number of credits taken, and other relevant attributes, using the Naive Bayes method. Naive Bayes was chosen because of its ability to classify data efficiently and accurately, even though the data used has noise or inconsistency. The research process begins with collecting student data from the university database, and then data cleaning is carried out to ensure the quality of the data used. Next, the data is processed and classified using the Naive Bayes algorithm in Weka software to predict graduation status based on academic parameters. The results show that the Naive Bayes method is able to produce quite high accuracy in predicting student graduation, with accuracy values ranging from 75% to more than 85% depending on parameter selection and data cleaning. GPA is the most influential attribute on the prediction results, while other attributes such as class activity and organizational experience also contribute, although not as much as GPA. These findings provide important insights for the campus in designing more effective academic coaching and planning programs and can be a reference in the development of data mining-based decision support systems to improve the quality of computer science graduates.

Keywords: Sentiment Analysis, Student Graduation, Naive Bayes Method, Academic, Computer Science.

Citation: Kalfida Eka Wati Siregar, Wily Supi Ramadani, Anggi Jelita Sitepu, Ulfi Muzayyanah Fadil, & Mhd. Furqan. (2025). Sentiment analysis of Faculty of Science and Technology students' satisfaction with the 2024 graduation using the Naïve Bayes method. *Iota*, 5(2). https://doi.org/10.31763/iota.v5i2.940

Academic Editor: Adi, P.D.P

Received: Maret 12, 2025

Accepted: April 23, 2025

Published: May 30, 2025

1. Introduction

In today's digital age, students often express their opinions and experiences through online survey platforms, social media, and academic forums. These opinions, if properly analyzed, can provide valuable insights into student perceptions of faculty services. One approach that is widely used to extract meaning from such opinion data is sentiment analysis[1]. The community's need for information and education, especially in the world of education, is an important thing that must be fulfilled. Educational institutions in Indonesia, especially universities, are one of the institutions that provide services to the community, especially in academic services. The quality of service is largely determined by the satisfaction of all stakeholders involved in the running of the college. Within the scope of higher education, these stakeholders include lecturers, educators, staff, industry, partners, students, parents of students, and the surrounding community.

Competition between universities in providing the best services requires educational institutions to continuously improve the quality of their services so as not to lag behind their competitors. The level of student satisfaction is a measure of the excellence of a college, this is a guarantee of quality which is the top priority for each college. If students are satisfied with the services provided, then the institution is considered to have good service. On the other hand, if students are not satisfied with the services provided, then the institution will receive a poor assessment [4,5,6].

One effective approach to exploring these perceptions is through sentiment analysis, which is an automated process to identify and categorize opinions expressed in text form. With this method, data in the form of comments or responses from students can be processed to determine whether the sentiment conveyed is positive, negative, or neutral. In this research, the Naive Bayes method is used, which is one of the probabilistic-based classification algorithms widely used in text analysis due to its simplicity and efficiency. This method works based on Bayes' Theorem with the assumption of independence between features (words in the text) and has proven effective in various text classification applications, including sentiment analysis [7,8,9].

The background of this research is based on the need for educational institutions to understand the patterns and factors that influence student graduation, especially in the field of computer science. So far, graduation evaluation has only focused on administrative data and academic performance, such as assignment grades, UTS, and UAS, thus paying less attention to student perceptions or sentiments towards the learning and evaluation process. Sentiment analysis can provide new insights into the challenges and obstacles faced by students so that they can be taken into consideration in making educational policies [2]. The formulation of the problem in this research is how to analyze the sentiment towards the graduation of computer science students based on academic data using the Naive Bayes method. In addition, this research also aims to identify the dominant factors that influence student sentiment towards the graduation process and see the extent to which the Naive Bayes method is able to provide accurate and relevant analysis results [3].

The research objective is to develop a sentiment analysis model that can classify the opinions of computer science students toward graduation based on academic data, using the Naive Bayes method [10,11,12]. This model is expected to be able to provide more in-depth information about student perceptions of the evaluation and graduation process so that it can be used as material for evaluating and improving the education system in higher education. This research is expected to contribute to the development of a higher education system that is more responsive to student needs and perceptions, as well as supporting data-based decision-making to improve the quality of computer science graduates. Analysis of academic-based graduation sentiment with the Naive Bayes method [16,17,18] is expected to provide accurate and useful results for educational institutions.

2. Method

The research method consists of several stages, i.e., Data Collection, Data Labelling, Data Preprocessing, and Data Weighting, then continued with Naive Bayes Algorithm Implementation, and Result Analysis [13,14,15].

2.1 Research Stages

2.1.1 Data Collection

Data Crawling: Collecting data in the form of comments or reviews related to the graduation of computer science students from social media, forums, or surveys.

2.1.2 Data Labelling

o Labeling: Labeling each data, for example: positive, negative, or neutral.

2.1.3 Data Preprocessing

- Data Cleaning: Removing symbols, links, and irrelevant characters.
- Case Folding: Converts all letters to lowercase.
- Tokenization: Breaks the text into individual words.
- Filtering/Stopword Removal: Removes unnecessary words (stopwords).
- Stemming: Converting words into their base form.

- Normalization: Converting non-standard words into standardized ones.

Table 1. Preprocessing Stage

Preprocessing Stage	Description
Case Folding	Change all letters to lowercase
Tokenization	Break text into separate words.
Stopword Removal	Delete unnecessary words
Stemming	Change words to their base form.
Normalization	Change nonstandard words to standard.

2.1.4 Data Weighting

- TF-IDF: Converts text data into numeric vectors.

2.1.5 Naive Bayes Algorithm Implementation

- Model Training: Building a classification model using the Naive Bayes algorithm.
- Model Evaluation: Testing the accuracy of the model with testing data.

The Naïve Bayes formula is shown in equation 1. Where $p(c|d)$ is the probability of c class which given by document d. and $P(d|c)$ is a Probability of document d on the class c, Moreover, $P(c)$ is a probability of class c, and $P(d)$ is a probability of document d.

$$p(c|d) = \frac{P(d|c).P(c)}{P(d)} \quad (1)$$

2.1.6 Result Analysis

- Interpretation: View sentiment results and evaluate model performance.

Table 2. Data after Preprocessing

Original Text	Text After Preprocessing	Label
"*Saya sangat BAHAGIA dengan kelulusan saya*"	"*Bahagia kelulusan*"	Positive
"*Saya kecewa dengan hasil kelulusan*"	"*kecewa hasil kelulusan*"	Negative
"*Kelulusan biasa saja*"	"*biasa*"	Neutral

Table 3. Summary of Research Stages

Step	Short Description
Data Collection	Crawling data from social media or surveys
Labeling	Labeling the sentiment
Preprocessing	Letter folding, tokenization, stopword, stemming, normalization
Weighting	TF-IDF for vectorization
Naive Bayes Model	Model training and evaluation
Analysis of Results	Interpretation of sentiment results and model accuracy

This research begins with data collection in the form of opinions or comments from computer science students related to graduation and academic aspects, for example through online questionnaires or interviews. The data collected includes student responses to the graduation process, academic factors such as GPA, credit load, and other academic experiences. The raw data obtained is then tidied up through a preprocessing stage to remove irrelevant data, such as unimportant words or duplication so that it is ready for analysis.

Once the data is clean, sentiment labeling is performed based on polarity, i.e. positive, negative, or neutral. Then, the data is divided into two parts: training data and testing data. The training data is used to build the Naive Bayes classification model, while the test data is used to test the accuracy of the model. At this stage, important features of student opinions are extracted to strengthen the sentiment classification process.

Furthermore, after the Naive Bayes model is built, testing is conducted using test data. The sentiment classification results are evaluated using metrics such as accuracy, precision, recall, and f1-score. This evaluation aims to determine the extent to which the model is able to predict student sentiment toward graduation and academic aspects. The evaluation results can be used as a basis for recommendations for the campus to improve the graduation process and academic services for computer science students. This research method ensures that the data used is clean and ready for modeling so that the results of the sentiment analysis of computer science student graduation can be more accurate and relevant using the Naive Bayes algorithm. Table 4 shows the Clawring Result Data [19,20].

3. Result and Discussion

Table 4. Data Clawring Result

Index	Timestamp	Email	Full Name	NIM	Cumulative Grade Point Average (IPK)	Number of credits	Study Program	Sentiments	Critique/Suggestion
0	2025-05-28, 11:37:35, 659000	kalfidaekawatisiregar@gmail.com	Kalfida Ekawati Siregar	701223101	NaN	NaN	Physics	Very Satisfactory	Verification of graduation data is done quickly and accurately, reducing student stress.
1	2025-05-28, 11:44:07, 545000	kalfidaekawatisiregar@gmail.com	Kalfida Ekawati Siregar	701202231	3.5	150.0	Physics	Satisfied	Supporting facilities for the graduation process are sufficiently available.
2	2025-05-28, 11:49:19, 360000	husensyah921@gmail.com	Hesein	701203864	3.54	153.0	Biology	Not Satisfied	Graduation experience gives a positive and satisfying impression to students.
3	2025-05-28, 11:52:54, 841000	riansyahrusma@gmail.com	Rianto	701203546	3.7	144.0	Computer Science	Satisfied	The graduation process helps students to continue their studies or enter the workforce easily.

Index	Timestamp	Email	Full Name	NIM	Cumulative Grade Point Average (IPK)	Number of credits	Study Program	Sentiments	Critique/Suggestion
4	2025-05-28, 12:04:34, 933000	delviaaulia09@gmail.com	Delvi Aulia	12210325723	3.8	132.0	Computer Science	Satisfied	Graduation administration services are friendly and attentive

Furthermore, Figure 1 is an example of criticism and suggestions. From the Criticism and Suggestion data inputted by users, these data can be analyzed in terms of positive and negative sentences from the Faculty of Science and Technology Graduation in 2024. These generated sentences are analyzed using the Naive Bayes strategy and method.

	Kritik_Saran
1	[Fasilitas, pendukung, untuk, proses, kelulusa...
2	[Pengalaman, kelulusan, memberikan, kesan, pos...
3	[Proses, kelulusan, membantu, mahasiswa, dalam...
4	[Pelayanan, administrasi, kelulusan, sangat, r...
5	[Kelulusan, menjadi, momen, yang, menyenangkan...
6	[Prosedur, kelulusan, terlalu, rumit, dan, mem...
7	[Tidak, ada, sosialisasi, yang, memadai, terka...
8	[Kelulusan, terkadang, diumumkan, secara, mend...
9	[Banyak, kesalahan, administrasi, dalam, dokum...
10	[Pengumuman, kelulusan, kurang, profesional, d...
11	[Mahasiswa, sering, dipersulit, dengan, persya...
13	[Proses, verifikasi, data, kelulusan, sangat, ...
14	[Kelulusan, terasa, tidak, adil, karena, adany...

Figure 1. Comments and Suggestions

Moreover, Figure 2 shows the Tokenized Result of the input from the Tokenization Analysis, which is shown with Index 0:"*fasilitasdukunguntukanproseskelulusesediadengnapada*", This token is related: facilities, support, process, and graduation, with the meaning that supporting facilities for the graduation process are well available. Index 1: "*alamkelulusberrikesanpositifanmuasbagimahasiswa*", Related tokens: nature, graduation, positive impression, satisfied, students, with Meaning: The graduation atmosphere gives a positive and satisfying impression to students. Index 2: "*proseskelulusbantumahasisdalamianjutstudiataupasukaniakerjadenganmudah*", Related tokens: process, graduation, aid, student, continuing studies, work, Meaning: The graduation process helps students continue their studies or find jobs easily. Index 3: "*layanadministrsikelulusangatramahdanpenuhperhati*", Related tokens: service,

administration, graduation, friendly, attentive, Meaning: Graduation administration services are friendly and attentive, Index 4: "*kelulusjadimonenyangsenangdanbangga berkatkordinasiyangbaik*". Related tokens: graduation, moment, happy, proud, good coordination. Meaning: Graduation is a happy and proud moment thanks to good coordination. Observation: All tokens show positive sentiments regarding student satisfaction. Main focus on aspects: facilities, graduation process, administrative services, and coordination. This data can be used for sentiment classification of "Satisfied" vs "Dissatisfied".

index	Kritik_Saran
0	fasilitasdukunguntukproseskelulussediadenganpada
1	alamkelulusberikesanpositifdanmuasbagimahasiswa
2	proseskelulusbantumahasiswadalamlanjutstudiataupasukduniakerjadenganmudah
3	layanadministrasikelulussangatramahdanpenuhperhati
4	kelulusjadimomenyangsenangdanbanggaberkatkoordinasiyangbaik

Figure 2. Tokenized Result

```
Komentar Positif:
                                                                                                  Kritik_Saran sentiment
5                                    Kelulusan menjadi momen yang menyenangkan dan membanggakan berkat koordinasi yang baik   positif
6                                                 Prosedur kelulusan terlalu rumit dan memakan waktu lama gajelas, tai   positif
9                              Banyak kesalahan administrasi dalam dokumen kelulusan yang harus diperbaiki gabisa bagus croot   positif
14  Kelulusan terasa tidak adil karena adanya ketidakkonsistenan dalam penilaian dan prosedur, gajelas bjir benerin dikit lol   positif

Komentar Negatif:
                                                          Kritik_Saran sentiment
13  Proses verifikasi data kelulusan sangat lambat dan membuat stres mahasiswa   negatif
```

Figure 3. Voting result

Moreover, from Figure 3, it can be seen that the positive comments are 4 documents (index: 5, 6, 9, 14), and the negative comments are 1 document (index: 13), a total of 5 documents, a ratio of 80% Positive, and 20% Negative. Imbalance data 80% positive vs 20% negative, Risk of bias towards the majority class (positive), and Prior Probability P(Positive) =4/5 =0.8, P(Negative)=1/5=0.2
Current Data Quality: Poor - Significant mislabeling issues
Model Reliability: Low - High risk of biased predictions
Immediate Actions Needed:

- Fix data labels
- Add more negative samples
- Clean and standardize text
- Implement proper evaluation metrics

Potential After Fixes: Good - With proper preprocessing and balanced data, Naive Bayes can work well for this sentiment analysis task.

```
   Nama Lengkap                          cleaned sentiment
0   Andi Wijaya       pelayanan cepat dan bagus   positif
1  Budi Santoso         proses lambat dan ribet   negatif
3  Dewi Lestari  pelayanan memuaskan dan mantap   positif
```

Figure 4. Filtering Result

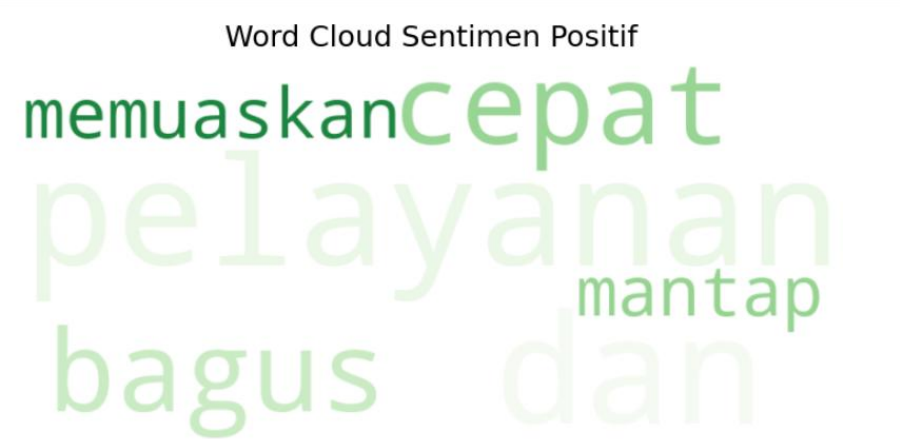

Figure 5. Positive Label Results

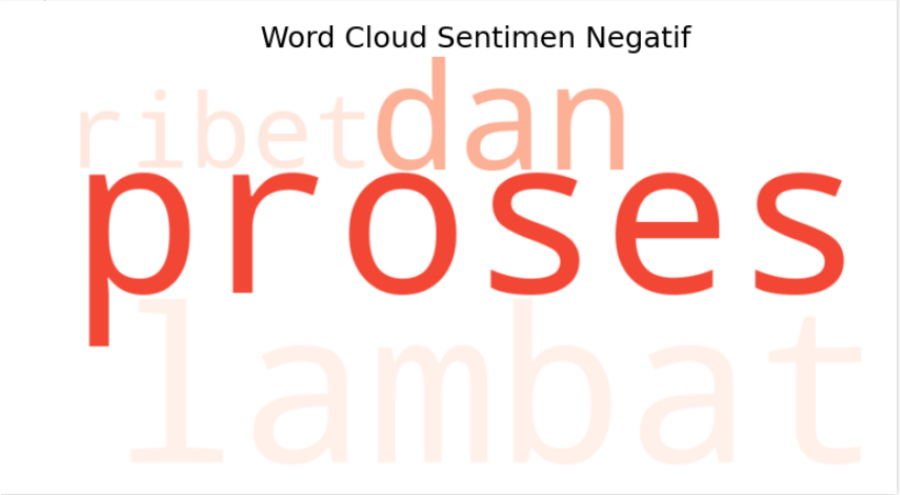

Figure 6. Negative Label Results

4. Conclusions

Sentiment analysis of graduation of Computer Science students based on academic data with the Naive Bayes method has proven effective in classifying opinions and predictions related to graduation. This method is able to utilize academic performance data such as assignment grades, UTS, UAS, and GPA as the main attributes that have a significant effect on classification results. Similar research results show that the use of the Naive Bayes algorithm produces a fairly high level of accuracy, even compared to other algorithms such as Random Forest and C4.5, especially in predicting student graduation. The application of sentiment analysis to student graduation data not only provides an overview of graduation trends but can also identify the most dominant factors in the academic process. For example, GPA and exam results proved to be the most influential attributes on the speed and predicate of graduation. Thus, this analysis provides important insights for educational institutions in understanding student academic dynamics and anticipating problems that may arise during the study process.

The Naive Bayes method was chosen due to its ability to handle relatively large data with fast computation time, as well as a stable accuracy rate in various academic prediction cases. In addition, the use of data mining and sentiment analysis allows universities to clean data from noise so that the classification results become more accurate and relevant to the actual conditions of students. Sentiment analysis of Computer Science student graduation with the Naive Bayes method can be the right solution to improve the quality of monitoring and evaluation of academic processes. Educational institutions can use the results of this analysis to make more targeted policies, such as providing special assistance for students who are at risk of delayed graduation or improving academic programs that support timely graduation.

Based on the research results, it is recommended that educational institutions make more use of sentiment analysis and data mining on an ongoing basis to monitor and evaluate student academic performance. In addition, it is necessary to develop a more adaptive prediction model by adding new attributes such as participation in organizations, internship experience, and students' psychological conditions to improve the accuracy and relevance of the analysis results. Institutions also need to pay attention to the data cleaning process so that the prediction and sentiment analysis results remain accurate and reliable in academic decision-making.

Acknowledgments: Thanks to all colleagues of the State Islamic University of North Sumatra, Medan, North Sumatra who have helped in the completion of this research, hopefully this research can continue to develop more specifically, so that it can contribute to the development of science in Indonesia.

Author contributions: The authors are responsible for building Conceptualization, Methodology, analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision of project administration, funding acquisition, and have read and agreed to the published version of the manuscript.

Funding: The study was conducted without any financial support from external sources.

Availability of data and Materials: All data are available from the authors.

Conflicts of Interest: The authors declare no conflict of interest.
Additional Information: No Additional Information from the authors.

References

[1] E. Miana, A. Ernamia, A. Herliana, A. R. Sanjaya, and A. R. Sanjaya, "Analisis Sentimen Kuliah Daring Dengan Algoritma Naïve Bayes, K-NN dan," vol. 4, no. 1, pp. 70–80, 2022.
[2] P. Metode, C. Algoritma, and D. A. N. Naïve, "Perbandingan metode algoritma c4.5 dan naïve bayes untuk memprediksi kelulusan mahasiswa," vol. 4, no. 1, pp. 11–22, 2022.
[3] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," 2018.
[4] S. Ufia, A. D. Nugroho, and T. Wahjoedi, "Meningkatkan Kompetensi Mahasiswa melalui Program Magang Sebagai Upaya Peningkatan Hard Skill dan Soft Skill," J. Knowl. Collab., vol. 1, no. 2, pp. 39–47, 2024, doi: 10.59613/97dmmj73.
[5] K. R. Diska and K. Budayawan, "Sistem Informasi Prediksi Kelulusan Menggunakan Metode Naive Bayes Classifer (Studi Kasus : Prodi Pendidikan Teknik Informatika)," vol. 7, pp. 936–943, 2023.
[6] K. Berdasarkan, D. Mahasiswa, U. Lancang, A. Zamsuri, and D. Muladi, "Perbandingan Akurasi Algoritma C4 . 5 Dan Naïve Bayes Terhadap Klasifikasi Kelulusan Berdasarkan Data Mahasiswa Universitas Lancang Kuning," no. x, pp. 398–406, 2015.
[7] J. Ilmiah, I. Administrasi, U. Islam, N. Sunan, and G. Djati, "Peran Penggunaan Teknologi Informasi Dan Komunikasi Dalam Budaya Organisasi Anisya Putri Andita, Ningrum Fauziah Yusuf," vol. 11, no. 2, 2021.
[8] S. I. Rachmayanti and M. Gufron, "Analisis Faktor yang Mengambat dalam Penanaman Pendidikan Karakter Disiplin pada Siswa di SDN 02 Serut," Insprasi J. Ilmu-Ilmu Sos., vol. 16, no. 2, pp. 124–131, 2019, [Online]. Available: https://jurnal.stkippgritulungagung.ac.id/index.php/inspirasi/article/view/1427

[9] M. F. Hakim and S. Saputra, "Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Analisis Prediksi Kelulusan Mahasiswa Universitas Dinamika Bangsa Menggunakan Metode Naïve Bayes Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)," vol. 5, no. April, 2025, doi: 10.33998/jakakom.v5i1.

[10] V. No and R. K. Ramli, "Edumatic : Jurnal Pendidikan Informatika Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes," vol. 6, no. 1, pp. 1–10, 2022, doi: 10.29408/edumatic.v6i1.4789.

[11] J. N. C. Gunawan, Y. Hadi, and N. K. Putrianto, "Penerapan System Development Life Cycle untuk Perancangan Sistem Pemenuhan Material pada Proses Produksi Minibus (Studi Kasus pada Departemen Perlengkapan Karoseri PT X)," J. Tekst. J. Keilmuan dan Apl. Bid. Tekst. dan Manaj. Ind., vol. 7, no. 1, pp. 69–80, 2024, doi: 10.59432/jute.v7i1.96.

[12] M. Kurnia DN and A. Wahyu Gunawan P, "Pengaruh Work-Life Balance Terhadap Employee Performance Yang Dimoderasi Oleh Peran Dari Organizational Support Dan Job Burnout Di Pt. Pertamina Hulu Rokan," Journal of Comprehensive Science (JCS), vol. 2, no. 5. pp. 1254–1262, 2023. doi: 10.59188/jcs.v2i5.343.

[13] A. Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," Zo. J. Sist. Inf., vol. 5, no. 1, pp. 59–70, 2023.

[14] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter," SMATIKA J. STIKI Inform. J., vol. 10, no. 02, pp. 71–76, 2020.

[15] B. Ramadhani, R. R. Suryono, and K. Kunci, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," vol, vol. 8, pp. 714–725, 2024.

[16] A. Sholihat, F. Bei, R. Ainaya, F. Sembiring and A. Lattu, "Twitter Tweet: Sentiment Analysis on Illegal Investment using Naive Bayes Algorithm," 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED), Sukabumi, Indonesia, 2022, pp. 1-5, doi: 10.1109/ICCED56140.2022.10010383.

[17] E. Hairani and I. Nurpalah, "Sentiment Analysis of Muslim Pro "A1-Quran Adzan" on Google Play Using Multinomial Naïve Bayes," 2024 12th International Conference on Cyber and IT Service Management (CITSM), Batam, Indonesia, 2024, pp. 1-5, doi: 10.1109/CITSM64103.2024.10775862.

[18] Y. Durachman, S. J. Putra, H. Nanang and H. T. Sukmana, "Analysis Sentiment of Public Opinion on Social Media Using Naïve Bayes and TF-IDF Algorithms," 2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT), Tangerang, Indonesia, 2024, pp. 1-6, doi: 10.1109/ICCIT62134.2024.10701191.

[19] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.

[20] N. S. Kustanto, I. Nurma Yulita and I. Sarathan, "Sentiment Analysis of Indonesia's National Health Insurance Mobile Application using Naïve Bayes Algorithm," 2021 International Conference on Artificial Intelligence and Big Data Analytics, Bandung, Indonesia, 2021, pp. 38-42, doi: 10.1109/ICAIBDA53487.2021.9689726.