Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi
Vol.9 No.2 July 2024, P-ISSN : 2502-3470, E-ISSN : 2581-0367

Sentiment Analysis of Student Complaint Text for Finding Context

Anisa Dzulkarnain[1], Berlian Rahmy Lidiawaty[2], Sri Hidayati[3], Rafi Andi Hidayah[4], Arini Pramesta Setyaningtitah[5]

[1,2,3,4,5]*Departement of Information System, School of Industrial Engineering, Telkom University, Indonesia*

[1]anisadzulkarnain@telkomuniversity.ac.id(*)

[2,3][berlianerel, srihidayatisi]@telkomuniversity.ac.id, [4,5][raprapandi, arinipramesta]@student.telkomuniversity.ac.id

Received: 2024-01-24; Accepted: 2024-05-17; Published: 2024-05-03

*Abstract***—** Students are one of the users of services provided by institutions. Student complaint questionnaires were distributed to find out various complaints related to institutional services. The problem is how to interpret student complaints so that institutional services can be improved according to what is needed. Therefore, this research aims to analyze the sentiment of student complaints. Apart form that, this research also aims to find the context of each student's complaint regarding institutional services. This research uses sentiment analysis with Indonesian text. The method starts from data collection, data cleaning and labelling of complaints, pre-processing of complaint text, term frequency (TF) method used to extract the content, and accuracy measurement. The labeling process was carried out twice to compare the accuracy, precision, recall, and f1-score of the models. Based on the results of the sentiment analysis of student complaints, the accuracy rate reached 76.1%. Additionally, a precision of 65.2% and recall of 85.2% indicate that labeling is more balanced, although there is still room for improvement.

Keywords— Sentiment Analysis; Text Mining; Classification; Student Complaint; Institutional Service.

I. Introduction

Higher education is an educational institution that provides services to students to achieve academic goals and personal development. As student service institutions, tertiary institutions are responsible for providing a conducive environment for students to learn, develop, and participate in academic and non-academic activities. Student complaints against institutional services can be divided into three sectors[1]. The first sector is the student's complaints about academic services, research, and community services. The second sector is the students' complaints about human resources, information and technology (IT) facilities and infrastructure, finance, logistics, and asset management. The third sector is student's complaints about student admissions, student affairs, and scholarships.

The main problem in understanding a student's complaint is how to interpret it without bias [2], [3], [4]. Sentiment analysis of textual data, including reviews and complaints, has been advanced to provide considerations for decision-making at the managerial level [5]. Therefore, this paper's objective is to analyze student's complaint text. This research not only analyzes the complaint's sentiment but also tries to find the context that will be used to understand why students tend to send their complaints. In order to obtain more comprehensive observational outcomes, this study employed a case study approach within a private higher education institution in Indonesia. Consequently, the data collected consisted of student complaints expressed in Bahasa Indonesia. As a result, text-processing techniques specific to the Indonesian language exhibit distinct characteristics compared to English [6].

After conducting a text mining process in Bahasa Indonesia, this study proceeded to ascertain the context of student complaint texts. Context pertains to the factor that imparts information within a given text[7], [8]. Fahrudin, Buliali, and Fatichah previously researched student reviews regarding campus facilities. This research created a multi-label text classification framework to categorize student review texts related to Telkom University facilities into several classes, such as HR issues, relations, and related agent institutions. This classification uses the word embedding method by matching the baf-of-words with the input review text[9].

Other research related to sentiment analysis of lecturer evaluation review texts by university students was also conducted by Amrusian, Widayat, and Wirawan. This research uses the Long Short-Term Memory (LSTM) method to classify reviews as having a positive or negative meaning. Their results obtained an accuracy of up to 91.08%[10]. In other studies, sentiment analysis has been used to classify a text as having a positive or negative sentiment value and is usually used for decision-making in this case. Positive sentiment values usually indicate that a text has positive nuances that indicate the meaning of goodness in a context. Similar to previous research conducted by Guerreiro and Rita, which utilized the positive sentiment of text reviews to serve as a basic for hotel owners, what facilities need to be improved[11]. Negative sentiments are usually used to identify weaknesses in the context being discussed. An example of research conducted by Park, et al., which utilized negative sentiment analysis to make decisions about which public transportation places should implement more health protocol policies during the Covid 19 pandemic [12]. In this study, a sentiment analysis of student complaint texts was conducted regarding institutional services. The purpose of sentiment analysis in the context of this research is to identifiy and extract subjective information from student feedback and complaints about institutional services to understand student satisfaction.

II. Research Methodology

DOI : https://doi.org/10.25139/inform.v9i2.7743

This chapter explains the methods that are used in this paper. It starts with data collection, data cleaning, complaint labeling, complaint text pre-processing, context extraction, and accuracy measurement. The research stages are shown in Figure 1.

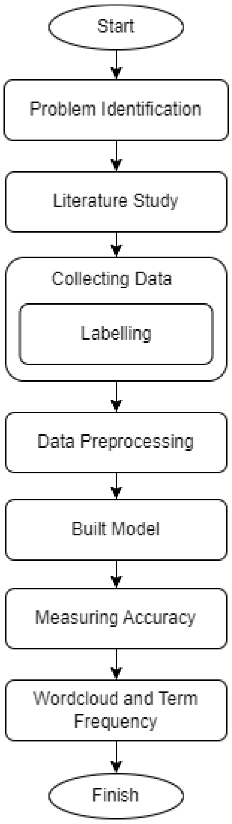


Figure 1. Research Stages

A. Data Collection

At the data collection stage, this study collected student's complaints about their institution. The institution, as the study case in this study, is a private institute in Surabaya. There are 2.084 student complaint texts collected. The sample data that was used in this study can be seen in Table I.

TABLE I
SAMPLE DATA OF STUDENT'S COMPLAINT TEXTS

Faculty	Department	Student's Complaint	Human Sentiment Label		
			ul_1	ul_2	ul_3
f_abc	dept_a	"*Sebaiknya akun sosmed ******** juga memberikan info beasiswa dari luar dan info dari program kampus merdeka*." Meaning: "It is best if the *campus*'s social media account also provides information on scholarships from outside and information from the independent campus program."	Positive	Negative	Negative
f_abc	dept_b	"*cukup baik*" meaning: "good enough"	Positive	Positive	Positive
f_abc	dept_c	"*parkirannya tolong diperluas*" Meaning: "need more parking area."	Negative	Negative	Negative

In Table I, there are three attributes from the original data: faculty, program study, and student complaints. The faculty and program study provides information about the student. The student's complaint is the string text that the student sent to the institute's quality assurance department. Furthermore, there is a 'human sentiment label' attribute used to store the sentiment labels assigned by humans to the complaint texts submitted by the students. The reason there are 3 User Labelers (ul_x) in Table 1 is that there are three people who label every student's complaint.

B. Data Cleaning and Complaint Labelling

Prior to obtaining the data for this study, a total of 2.341 rows of student complaints were initially detected. However, during this stage, it was necessary to eliminate rows that met specific criteria, such as those containing only symbols, question marks, or other specific criteria, such as those containing only symbols, question marks, or other specific patterns. The data was manually cleaned by removing rows that fulfilled these criteria. Following this process, 2.077 rows of data remained, which were ready for the subsequent labeling phase.

Following the normalization of the data, the subsequent step involves assigning labels to each student complaint. This labelling process was conducted by three users, as illustrated in Table 1. The intentional use of an odd number of labelers aims to prevent potential tie situations where two classes have an equal number of voters[13]. Moreover, there are three options to label a student's complaint: user labeler 1 (ul_1), user labeler

DOI : https://doi.org/10.25139/inform.v9i2.7743

2 (ul_2), and user labeler (ul_3). Pos stands for positive, and neg stands for negative. Once the three labelers have assigned their labels, the rules for determining the value of rules outlined in Table II are based on the calculation of the pos or neg sentiments rather than the specific order in which the labelers assign the sentiment values.

TABEL II
RULE TO DETERMINE HUMAN SENTIMENT VALUE

ul_1	ul_2	ul_3	human_sentiment	Rule
pos	pos	neg	positive	If there are at least two positives, then it's positive
neg	neg	pos	negative	If there are at least two negatives, then it's a negative

C. Complaint Text Pre-processing

After the data is collected and normalized, the next stage is pre-processing the student's complaint text. Data pre-processing aims to prepare raw data in a format that is more easily understood and processed by text-mining algorithms[14]. The pre-processing stages are explained in Figure 2. The text pre-processing in this study can be seen in Figure 2.

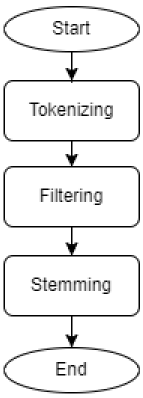


Figure 2. Data Pre-processing

Based on Figure 2, the text pre-processing started with case folding to lowercase and continued with punctuation and symbol removal. The second step is stopword removal or removing the words that commonly appear but do not have special meaning, usually containing conjunctions such as but, and then (in *Bahasa Indonesia*: *tetapi*, dan, *kemudian*) and/or any similar word[15], [16].

The third step is tokenizing or separating text into individual tokens using the NLTK library. Tokenization involves changing a sentence into a series of words or something that is often referred to as a token[17]. This tokenization aids in textual data analysis, for example, when creating a dictionary, editing certain words, calculating the frequency of words, or data such as out-of-vocabulary words (OOV)[18]. This diminishes the look method and decreases the capacity space required to store the information, which is the preference for utilizing tokenization[19]. The last text pre-processing is stemming or changing words to their root, like removing the prefix and suffix. The steaming process was conducted by implementing the Sastrawi library.

D. Context Extraction

The process of context extraction can be performed by calculating the term frequency (TF) value[8]. Context extraction by calculating the TF value is also commonly known as text feature selection[20]. The TF value of a word was calculated based on its frequency of occurrence across the entire corpus of the student's complaint text[21]. In the text pre-processing stage, preceding stop word removal was carried out in order to maximize the TF value computation[22]. This is because words appearing in stop words tend to possess higher TF values[8].

E. Accuracy Measurement

This study was classified as supervised learning since the student's complaint texts were initially labeled. To assess the accuracy of the sentiment analysis model, particularly in the context of institutional service, as investigated in this study, a confusion matrix is utilized to estimate accuracy, precision, and recall[23]. The confusion matrix typically consists of four values: true positive (TP), true negative (TN), false positive (FP), and false negative (FN)[24]. TP is obtained when the human_sentiment value in Table II is positive, and the system's sentiment prediction is also positive. TN is achieved when both the human sentiment and the system's prediction are negative. FP occurs when the system incorrectly predicts a positive sentiment while the human sentiment label is negative. FN is obtained when the system predicts a complaint as negative, but the human sentiment is positive. The accuracy, precision, and recall values were obtained from the results of the confusion matrix. If the confusion matrix results from the resulting model had an accuracy of less than 50%, a double-labelling process was performed. The double-labelling process was carried out to increase the accuracy of the resulting model. The scenario of increasing the accuracy values using double labelling, as done by Wang et al.[25]. Mapping these six accuracy values can be observed using a confusion matrix that is presented in Table III.

TABLE III
CONFUSION MATRIX IN THIS RESEARCH

		system_sentiment	
		positive	**negative**
human_sentiment	**positive**	TP	FN
	negative	FP	FN

The accuracy values derived from the comparison between human_sentiment labels and system predictions in Table 3 can be employed to compute various accuracy metrics, including accuracy (1), recall (2), precision (3), and F1 score (4). A high accuracy value indicates a better sentiment classification model[26].

$$Accuracy\ (a) = \frac{TP+TN}{TP+FN+FP+FN} \tag{1}$$

$$Precision\ (p) = \frac{TP}{TP+FP} \tag{2}$$

$$Recall\ (r) = \frac{TP}{TP+FN} \tag{3}$$

$$F1\text{-}score = \frac{2.p.r}{p+r} \tag{4}$$

DOI : https://doi.org/10.25139/inform.v9i2.7743

III. RESULT AND DISCUSSION

The data labeling stage is carried out after the data cleaning process. The labeling process was done manually on 2077 data lines, resulting in 827 positives and 1250 negatives, based on the rules in Table II. After data pre-processing, the sentiment classification process is carried out. The system_sentiment results are then compared with the human_sentiment results. The system_sentiment results consist of 1049 positive and 1028 negative. The percentage of human_sentiment and system_sentiment results is shown in Figure 3. The confusion matrix human_sentiment and system_sentiment can be seen in Table IV.

(a) The Percentage of Human_Sentiment

(b) The Percentage of System_Sentiment

Figure 3. Positive and Negative Sentiment Analysis Results

TABEL IV
CONFUSION MATRIX IN THIS RESEARCH

		system_sentiment	
		positive	**negative**
human sentiment	**positive**	704	122
	negative	375	876

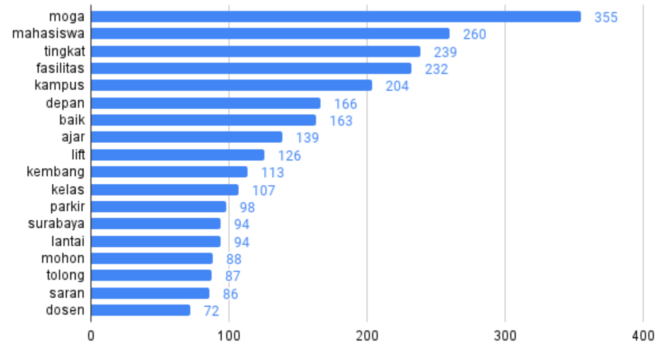
In this research, a double labelling process was conducted. This has an impact on increasing the accuracy, precision, and recall values between the first and second labeling. The comparison between the first and second labelling can be seen in Table 5.

TABEL V
COMPARISON BETWEEN FIRST LABELING AND SECOND LABELING

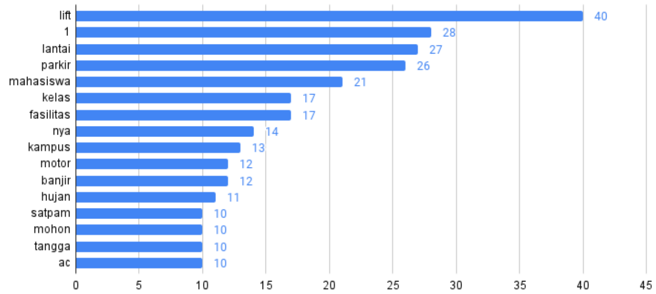
Labelling Process	Accuracy	Precision	Recall	F1-score
First labelling	32.4%	40.7%	96.9%	57.3%
Second labelling	76.1%	65.2%	85.2%	73.8%

In this first labeling process, the accuracy rate only reached 32.4%, indicating that most of the labels provided were inaccurate. Although the precision rate of 40.7% indicates that when a label is given, it is most likely correct, the very high recall rate of 96,9% indicates that many instances that should be labeled are not labeled. As a result, the low F1-score value of 57,3% illustrates the subtlety between precision and recall in this labeling.

Meanwhile, in the second labeling process, there was a significant increase in accuracy level, reaching 76.1%. This shows that the labels provided are more accurate than the first labeling. In addition, the precision of 65.2% and recall of 85.2% indicate that labeling is more balanced, although there is still room for improvement. The result is a higher F1-score, reaching 73.8%, illustrating better label quality in the second labeling. This significant improvement shows the importance of relabeling to improve data quality. To provide more detailed information on our data, we visualized each sentiment category using word clouds in Figure 6. We obtained these word clouds from the corpus positive and corpus negative.



(a) Frequency of Positive Sentiment



(b) Frequency of Negative Sentiment

Figure 4. Frequency Data Visualization of Sentiment Analysis Results

Figure 4 (a) Ais name a visualization of positive sentiment frequency data, which contains eighteen positive words that often appear in student complaints. The number of frequencies of the word "moga" is 355 times discussed in positive sentiment, followed by the word "*mahasiswa*" 260 times, the word "tingkat" 239 times, the word "*fasilitas*" 232 times, and the word "kampus" 204 times. Figure 4 (b) is a visualization of negative sentiment frequency data where eighteen negative words often appear in student complaints. The number of frequencies of the word "lift" is 40 times discussed in negative sentiment, followed by the word "1" 28 times, the word "*lantai*" 27 times, the word "*parkir*" 26 times *and* the word "*mahasiswa*" 21 times.

(a) Positive Sentiment

(a) Negative Sentiment

Figure 5. Visualization of the Most Frequently Used Keyword

IV. CONCLUSION

DOI : https://doi.org/10.25139/inform.v9i2.7743

This research analyzes student complaint texts by comparing system_sentiment results with human_sentiment results. The labeling process is carried out twice to increase the accuracy, precision, recall, and f1-score values. The confusion matrix of this research resulted in true positives of as many as 704 complaints, true negatives of 876 complaints, false negatives of 122 complaints, and false positives of as many as 375 complaints. The accuracy value of the second labeling shows that 76.1% of the model is accurate. This is directly proportional to the F1-score value, which increased in the second labeling by 73.8%. The value obtained represents that the sentiment system results are better. Positive words that often appear consist of “moga”, “mahasiswa”, “tingkat”, “fasilitas”, and “kampus”. The most frequently occurring negative words consisted of "*lift*", 1, "*lantai*", "*parkir*", and "*mahasiswa*". Suggestions for improvement in further research can be classified into three sentiments, namely positive, negative, and neutral.

ACKNOWLEDGMENT

The research has been partially supported by the research institute and community service (LPPM) of Telkom University Surabaya.

REFERENCES

[1] D.Telkom, *Copy of SOTK - 2019*. Indonesia, 2019, pp. 1–26.
[2] H. Yin, X. Song, S. Yang, and J. Li, “Sentiment analysis and topic modeling for COVID-19 vaccine discussions,” *World Wide Web*, vol. 25, no. 3, pp. 1067–1083, May 2022, doi: 10.1007/s11280-022-01029-y.
[3] T. Ling, L. Jake, J. Adams, K. Osinski, X. Liu, and D. Friedland, “Interpretable machine learning text classification for clinical computed tomography reports – a case study of temporal bone fracture,” *Computer Methods and Programs in Biomedicine Update*, vol. 3, p. 100104, 2023, doi: https://doi.org/10.1016/j.cmpbup.2023.100104.
[4] A. Balayn, C. Lofi, and G. J. Houben, “Managing bias and unfairness in data for decision support: a survey of machine learning and data engineering approaches to identify and mitigate bias and unfairness within data management and analytics systems,” *VLDB Journal*, vol. 30, no. 5, pp. 739–768, Sep. 2021, doi: 10.1007/s00778-021-00671-8.
[5] Q. Wang, W. Zhang, J. Li, F. Mai, and Z. Ma, “Effect of Online Review Sentiment on Product Sales: The Moderating Role of Review Credibility Perception,” *Comput. Hum. Behav.*, vol. 133, no. C, Aug. 2022, doi: 10.1016/j.chb.2022.107272.
[6] B. R. Lidiawaty, M. E. Zulfaqor, O. Diyantara, and D. R. S. Dewi, “Keywords Generator From Paragraph Text Using Text Mining in Bahasa Indonesia,” in *2022 Interdisciplinary Research in Technology and Management (IRTM)*, 2022, pp. 1–4. doi: 10.1109/IRTM54583.2022.9791753.
[7] A. Carrera-Rivera, F. Larrinaga, and G. Lasa, “Context-awareness for the design of Smart-product service systems: Literature review,” *Comput Ind*, vol. 142, p. 103730, 2022, doi: https://doi.org/10.1016/j.compind.2022.103730.
[8] B. R. Lidiawaty, E. Suryani, and R. A. Vinarti, “Analyzing Traffic-Complaint Tweets Based on Time-Location Context Analysis to Develop A Traffic Urgency Model,” in *2023 8th International Conference on Business and Industrial Research (ICBIR)*, 2023, pp. 585–590. doi: 10.1109/ICBIR57571.2023.10147706.
[9] T. Fahrudin, J. L. Buliali, and C. Fatichah, “Ina-BWR: Indonesian bigram word rule for multi-label student complaints,” *Egyptian Informatics Journal*, vol. 20, no. 3, pp. 151–161, 2019, doi: https://doi.org/10.1016/j.eij.2019.03.001.
[10] M. A. Amrustian, W. Widayat, and A. M. Wirawan, “Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 535, Jan. 2022, doi: 10.30865/mib.v6i1.3527.
[11] J. Guerreiro and P. Rita, “How to predict explicit recommendations in online reviews using text mining and sentiment analysis,” *Journal of Hospitality and Tourism Management*, vol. 43, pp. 269–272, 2020, doi: https://doi.org/10.1016/j.jhtm.2019.07.001.
[12] J. Y. Park, E. Mistur, D. Kim, Y. Mo, and R. Hoefer, “Toward human-centric urban infrastructure: Text mining for social media data to identify the public perception of COVID-19 policy in transportation hubs,” *Sustain Cities Soc*, vol. 76, p. 103524, 2022, doi: https://doi.org/10.1016/j.scs.2021.103524.
[13] J. Salminen, A. M. Kamel, S.-G. Jung, and B. J. Jansen, “The Problem of Majority Voting in Crowdsourcing with Binary Classes,” 2021, doi: 10.18420/ecscw2021_n12.
[14] Muhammad Ilham Fadila, Hanafi, and Anggit Dwi Hartanto, “Sentiment Analysis of the Indonesian National Team in the 2020 AFF Cup Using Naïve Bayes and K-Nearest Neighbor Algorithms,” *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, pp. 1–6, Jan. 2023, doi: 10.25139/inform.v8i1.5222.
[15] D. Ramachandran and R. Parvathi, “Analysis of Twitter Specific Pre-processing Technique for Tweets,” *Procedia Comput Sci*, vol. 165, pp. 245–251, 2019, doi: https://doi.org/10.1016/j.procs.2020.01.083.
[16] D. Purwitasari, C. B. P. Putra, and A. B. Raharjo, “A stance dataset with aspect-based sentiment information from Indonesian COVID-19 vaccination-related tweets,” *Data Brief*, vol. 47, p. 108951, 2023, doi: https://doi.org/10.1016/j.dib.2023.108951.
[17] D. D. Nur Cahyo *et al.*, “Sentiment Analysis for IMDb Movie Review Using Support Vector Machine (SVM) Method,” *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, pp. 90–95, Mar. 2023, doi: 10.25139/inform.v8i2.5700.
[18] M. A. Borroto and F. Ricca, “SPARQL-QA-v2 system for Knowledge Base Question Answering,” *Expert Syst Appl*, vol. 229, p. 120383, 2023, doi: https://doi.org/10.1016/j.eswa.2023.120383.
[19] P. Y. Mahajan and D. P. Rana, “Text mining approach for the prediction of disease status from discharge summaries using CCBE and NEROA-CNN,” *Expert Syst Appl*, vol. 227, p. 120310, 2023, doi: https://doi.org/10.1016/j.eswa.2023.120310.
[20] P. Dhal and C. Azad, “A comprehensive survey on feature selection in the various fields of machine learning,” *Applied Intelligence*, vol. 52, Jul. 2021, doi: 10.1007/s10489-021-02550-9.
[21] J. Yoon and H. Son, “Need differences by treatment phases between patients with colorectal cancer and their caregivers: A text mining analysis,” *Asia Pac J Oncol Nurs*, vol. 9, no. 5, p. 100061, 2022, doi: https://doi.org/10.1016/j.apjon.2022.03.013.
[22] L. Cahyaningrum, A. Luthfiarta, and M. Rahayu, “Sentiment Analysis on the Impact of MBKM on Student Organizations Using Supervised Learning with Smote to Handle Data Imbalance,” *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, 2024, doi: 10.25139/inform.v9i1.7484.
[23] J. Pence, P. Farshadmanesh, J. Kim, C. Blake, and Z. Mohaghegh, “Data-theoretic approach for socio-technical risk analysis: Text mining licensee event reports of U.S. nuclear power plants,” *Saf Sci*, vol. 124, Apr. 2020, doi: 10.1016/j.ssci.2019.104574.
[24] H. K. Duan, M. A. Vasarhelyi, M. Codesso, and Z. Alzamil, “Enhancing the government accounting information systems using social media information: An application of text mining and machine learning,” *International Journal of Accounting Information Systems*, vol. 48, p. 100600, 2023, doi: https://doi.org/10.1016/j.accinf.2022.100600.
[25] Y. Wang, J. T. Zhou, and X. Song, “A RaaS model based on emotion analysis and double labeling applied to mobile terminal,” *IEEE Access*, vol. 6, pp. 70974–70982, 2018, doi: 10.1109/ACCESS.2018.2880738.
[26] M. N. Ashtiani and B. Raahemi, “News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review,” *Expert Syst Appl*, vol. 217, p. 119509, 2023, doi: https://doi.org/10.1016/j.eswa.2023.119509.