International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

Vol. 43 No. 2 March 2024, pp. 306-313

User Opinion Polarization on IPDN Jatinangor

Sirojul Alam, H.A Danang Rimbawa, Bambang Suharjo

Master of Cyber Defense Engineering, Faculty of Defense Science and Technology
The Republic of Indonesia Defense University
Bogor, Indonesia
sirojul.alam@tp.idu.ac.id

Abstract— The interaction of Indonesian society in the cyber environment is increasing alongside the growing ease of internet access. BPS data indicates a significant number of Indonesians regularly accessing the internet, exceeding 50% of the country's population. This interaction facilitates people to express their opinions, as observed on platforms like google.com for evaluating various places. IPDN Jatinangor is not exempt from public opinion. We gathered and conducted a thorough analysis of these opinions to understand the societal perception towards IPDN, particularly its main campus in Jatinangor. Our analysis employed polarization techniques based on the Indonesian language lexicon dictionary. The results of the analysis show that 37% of public opinions are positive, 29% are negative, and 34% are neutral.

Keywords—opinion mining; google; review; nlp;

I. Introduction

A reputable educational institution serves as a tangible manifestation of high-quality and competitive human capital. Its standards, encompassing facilities, teaching staff, students, graduates, and stakeholders, are subject to scrutiny in the increasingly digital landscape where online interactions proliferate. Statistical data from the Indonesian Central Statistics Agency (BPS), released in August 2023, underscores this trend, revealing a notable increase in internet usage among the Indonesian population, from 62.10% in 2021 to 66.48% in 2022 [1]. This surge reflects a growing technological literacy among Indonesians, highlighting the pervasive influence of internet technology in contemporary society.

As human interaction in the digital age, communications have transitioned from analog to digital platforms facilitated by the internet [2]. This shift has led to a proliferation of data, including videos, images, and text, generated through online interaction [3]. Textual data, exemplified by online reviews, provides valuable insights and evaluations from individuals regarding various subjects, serving as a benchmark for quality assessment [4]. User satisfaction, evident in online reviews, holds paramount importance as it offers insights into customer preferences and organizational performance [5]. These reviews serve as critical indicators of an organization’s reputation and success in creating positive impressions [6].

Sentiment analysis, also known as opinion mining, is a pivotal method within natural language processing (NLP) for extracting and categorizing opinions from textual data, such as reviews and tweets [7], into positive or negative sentiments [8]. This technique, utilized in diverse domains including marketing, customer service, and social media analysis, relies on techniques such as machine learning, data mining, artificial intelligence, and computational linguistics [9] [10].

[11] introduced a range of techniques and tools aimed at conducting sentiment analysis specifically tailored for social media platforms. This area of research has garnered significant attention, with numerous scholars exploring sentiment analysis across

different domains. For instance, [12] undertook the classification and characterization of tweet data, while [13] analyzed sentiments surrounding electronic money usage. Similarly, [10] focused on sentiment analysis related to Indonesia Immigration data, and [14] conducted sentiment analysis on tweets written in Arabic. Moreover, [15] contributed to the field by conducting studies on social issues incorporating sentiment analysis. These studies collectively underscore the versatility and applicability of sentiment analysis techniques across diverse contexts, as demonstrated by the insights provided in [11].

Opinion mining, employing methodologies such as machine learning, data mining, artificial intelligence, and computational linguistics, finds widespread application across diverse fields including marketing, customer service, social media analysis, and politics. The general procedure typically encompasses several key stages, including data acquisition, identification, extraction, classification, and evaluation, as delineated by [16] [17]. These fundamental steps form the basis for analyzing textual data, enabling the extraction of valuable insights regarding opinions and sentiments. Such insights serve as invaluable resources for informing decision-making processes and strategic initiatives across various industries and domains.

The Institute of Domestic Governance (IPDN) operates as a prestigious educational institution overseen by the Ministry of Home Affairs of The Republic of Indonesia, dedicated to enhancing human resource capabilities within the governance sector. With a sprawling presence across the nation, IPDN boasts 8 campuses strategically positioned in key locations. Among these, the main campus is situated in Jatinangor, Sumedang, West Java [18], serving as the epicenter of IPDN's educational endeavors. Alongside Jatinangor, Sumedang, IPDN extends its footprint to Jakarta, West Sumatra, West Kalimantan, North Sulawesi, South Sulawesi, West Nusa Tenggara, and Papua, ensuring widespread accessibility to its educational offerings.

IPDN's core mission revolves around the preparation of domestic governance cadres primed for deployment within local and central government structures. Renowned for its disciplined approach to education, IPDN instills a comprehensive skill set in its students, enabling them to navigate the complexities of governance effectively. Through a rigorous curriculum and practical training initiatives, IPDN cultivates a cadre of professionals who embody integrity, professionalism, and a holistic understanding of governance principles. These qualities are pivotal for shaping future leaders capable of driving positive change and fostering sustainable development across Indonesia's administrative landscape.

II. METHOD

Several methods, techniques, and tools utilized for data mining have been introduced by numerous researchers [11]. Furthermore, many have employed positive-negative polarity techniques [8] [19] and have utilized established computational algorithms such as support vector machines [20], logistic regression [21] [22], as well as naive Bayes [10] [23] [24] [25] [26] [14] [27] [15]. These algorithms have demonstrated their efficacy in achieving tasks with accuracy performance levels exceeding 80%.

This study adopts polarity techniques by leveraging the lexicon dictionary of Bahasa Indonesian. The Bahasa Indonesia lexicon dictionary serves as a comprehensive repository providing structured information on Indonesian language words [28]. It is constructed through a combination of resources, including the Kamus Besar Bahasa Indonesia (KBBI), Kateglo, and WordNet, catering to both human and machine accessibility. Continuously updated and expanded, the dictionary incorporates colloquial words and their standardized forms [29], along with synsets or sets of synonyms, for Indonesian words derived from monolingual lexical resources [30]. Furthermore, a unified WordNet Bahasa has been developed, serving as a valuable resource for studying lexical semantics in Malay languages, including Indonesia [31].

The study was conducted utilizing the Python programming language, employing polarization techniques for sentiment analysis. It commenced with data acquisition, involving the retrieval of review data from Google.com using the designed hashtag 'IPDN Jatinangor'. Prior to this stage, the 'instant data scraper' add-on for Google Chrome was installed to facilitate data collection, given that the research was conducted within the Chrome browser environment. The detailed procedural framework of the research endeavor is depicted in Fig. 1. below.

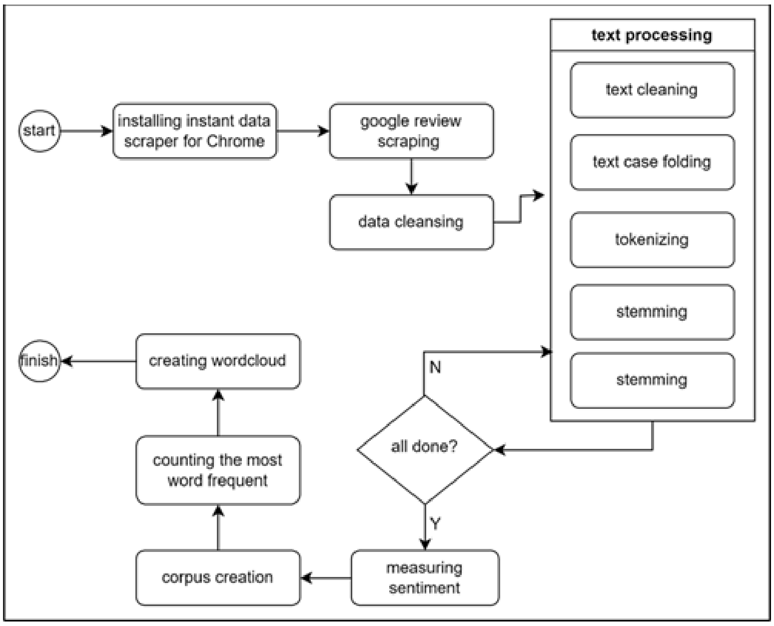

Fig. 1. Research steps

Fig. 1 illustrates the initial stages of data preprocessing following the successful retrieval of the dataset. The first task involved cleansing the dataset by removing entries with empty or null components, resulting in the elimination of numerous entries, leaving 375 review entries with substantive textual content. Subsequently, text preprocessing was conducted, comprising various essential steps. Text cleaning procedures were implemented to remove unwanted characters, duplicates, numbers, excessive spaces, and links. Additionally, case-folding was performed to standardize the text into lowercase format, ensuring consistency and eliminating unnecessary variations.

Standard Python libraries were utilized for these preprocessing tasks, followed by tokenization using the NLTK library to segment sentences into individual words. The subsequent filtering process retained only Indonesian words from the reviews, while stemming was employed to transform each word into its base form, devoid of prefixes, suffixes, and affixes. Each step underwent meticulous review to align with the research objectives, iterating until satisfactory outcomes were achieved before proceeding to sentiment analysis using the transformer library in Python.

The sentiment analysis phase aimed to gauge the overall sentiment expressed in the cleaned text data, providing valuable insights for subsequent analysis. A corpus was constructed from the cleaned text data, serving as a comprehensive collection of texts essential for natural language processing analysis [32]. This corpus facilitated the evaluation of models and determination of word intensity in user reviews [33]. Furthermore, the frequency of frequently occurring words was calculated and visualized in the form of a word cloud to offer a comprehensive overview of user opinion.

III. Result and Discussion

This study was carried out using the Google Colaboratory Notebook with the Python programming language. Several libraries were chosen to support the research goals, including those for text processing, data visualization, and data polarization calculation. We meticulously analyzed 375 text data entries extracted from Google reviews related to IPDN Jatinangor, following the cleansing process of initially obtained 1075 review entries. The results of our text processing efforts are presented in detail in Fig. 2 below.

review_clean	case_folded	tokenized	filtered	stemmed
Almamater tercinta	almamater tercinta	[almamater, tercinta]	[almamater, tercinta]	[almamater, cinta]
Suatu saat sy ingin masuk kesini	suatu saat sy ingin masuk kesini	[suatu, saat, sy, ingin, masuk, kesini]	[sy, masuk, kesini]	[sy, masuk, kesini]
IPDN	ipdn	[ipdn]	[ipdn]	[ipdn]
Tempat yang sangat sejuk ksatriaan terbaik di ...	tempat yang sangat sejuk ksatriaan terbaik di ...	[tempat, yang, sangat, sejuk, ksatriaan, terba...	[sejuk, ksatriaan, terbaik, indonesia]	[sejuk, ksatria, baik, indonesia]
Tuhan pasti menyertai Kakakku dlm perjuangan m...	tuhan pasti menyertai kakakku dlm perjuangan m...	[tuhan, pasti, menyertai, kakakku, dlm, perjua...	[tuhan, menyertai, kakakku, dlm, perjuangan, m...	[tuhan, serta, kakak, dlm, juang, raih, citaci...

Fig. 2. Text processing result

The dataset comprises several columns, each representing a distinct stage of text processing. In the ‘review_clean’ column, the text data undergoes a meticulous cleaning process, eliminating symbols, numbers, and superfluous special characters. Following this cleaning stage, the ‘case_folded’ column showcases words transformed into lowercase, ensuring uniformity in text representation. Subsequently, the ‘tokenized’ column displays the text data segmented into tokens, effectively breaking down sentences into individual words. After tokenization, the ‘filtered’ column retains solely Indonesian language words, filtering out non-Indonesian terms and enhancing data relevance. Finally, the ‘stemmed’ column contains text data converted into their base forms, facilitating streamlined analysis and interpretation.

To evaluate opinion polarization within the processed data, we employed sophisticated techniques leveraging the lexicon dictionary of the Indonesian language [34] [35] [36]. Utilizing this resource, sentiment analysis was conducted to gauge the extent of positive, negative, and neutral opinion polarization present. Through this analysis, we aimed to discern the prevailing sentiment within the dataset, providing valuable insights into user perceptions and sentiments. The results of the sentiment analysis are crucial for understanding the overall sentiment landscape surrounding the topic under investigation. Additionally, these insights can inform strategic decision-making processes aimed at enhancing user experience and satisfaction.

The application of opinion polarization techniques enables a nuanced understanding of the sentiment expressed within the text data. By leveraging the Indonesian language lexicon dictionary, we can accurately categorize opinions as positive, negative, or neutral. This categorization allows for a detailed examination of user sentiments, shedding light on areas of satisfaction and areas for improvement. The results of this analysis, detailing the opinion polarization, are visualized in Fig. 3. below.

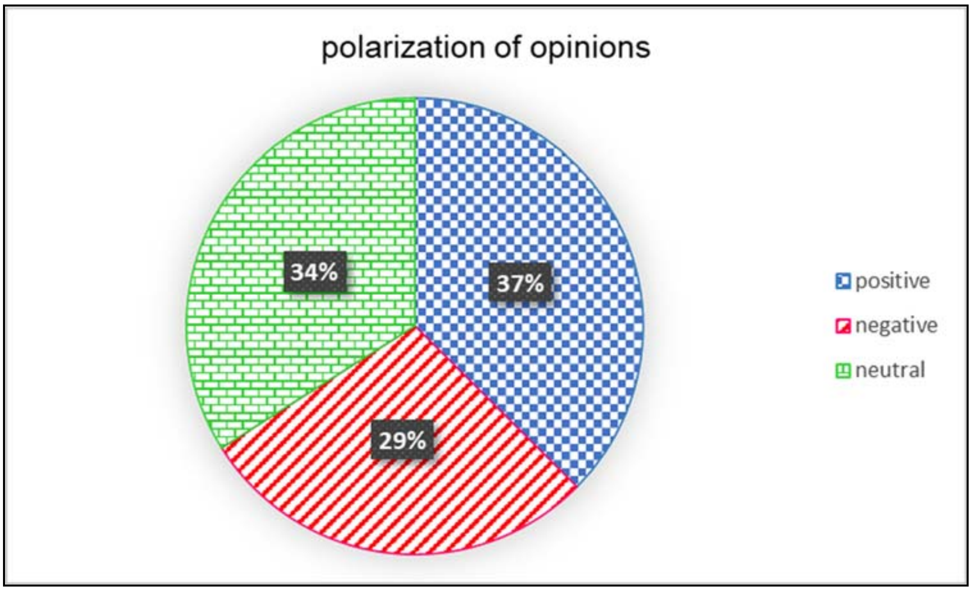

Fig. 3. User polarization

In Fig. 3, the sentiment polarization towards IPDN Jatinangor is depicted based on an analysis of Google reviews. The data reveals three distinct categories: positive, negative, and neutral. positive sentiment accounts for 37% of the total, negative sentiment for 29%, and neutral sentiment for 34%. Despite the slightly higher prevalence of neutral sentiment, the overall trend suggests a predominantly positive perception of IPDN Jatinangor among reviewers.

These findings highlight the importance of continuously monitoring and improving the institution's reputation. While the positive sentiment outweighs the negative, there remains room for enhancement to bolster public perception further. Analyzing the words frequently mentioned in reviews can offer valuable insights into areas of strength and areas requiring attention for IPDN Jatinangor's continued development and success.

Moving forward, leveraging these insights to address any areas of concern and further amplify positive aspects can contribute to the institution's ongoing improvement and reputation enhancement efforts. By proactively responding to feedback and implementing strategic initiatives based on data-driven insights, IPDN Jatinangor can continue striving towards excellence and ensuring its continued positive impact on stakeholders and the community. After identifying the types of polarization present, our subsequent step involved computing the most frequently occurring words in the reviews, as depicted in Fig. 4.

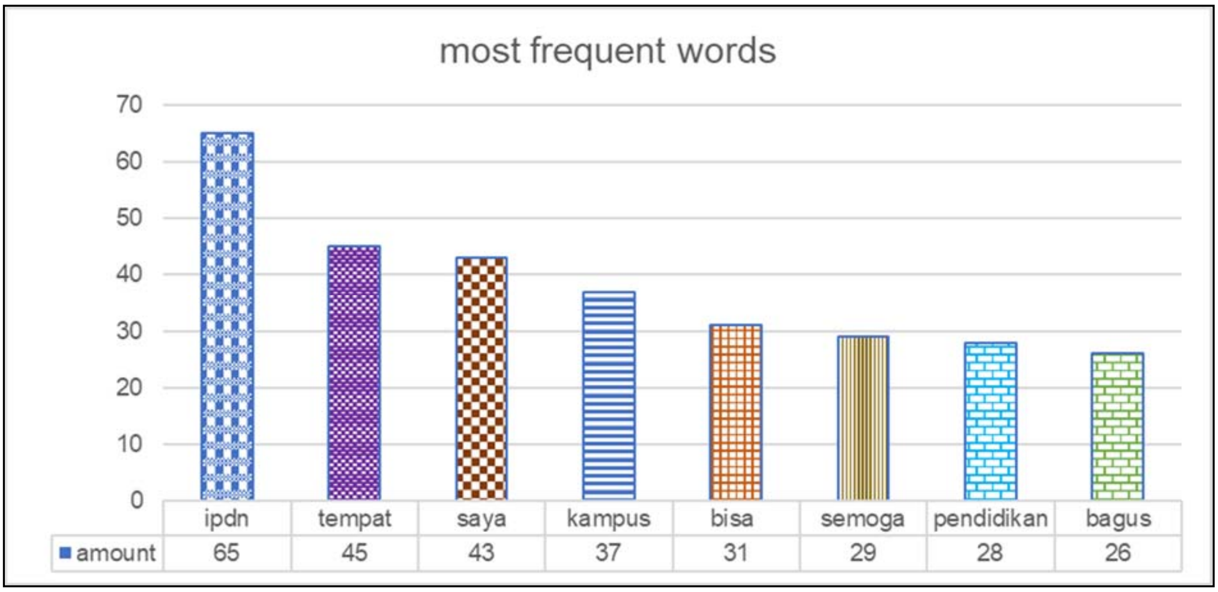

Fig. 4. Most frequent words

These words, depicted in Fig. 4, were sorted based on their frequency, with the x-axis representing the words that appeared most frequently in the reviews. Among them, the word 'ipdn' occurred 65 times, followed by 'tempat' (place) with 45 occurrences, 'saya' (I) with 43 occurrences, 'semoga' (hopefully) with 29 occurrences, and 'pendidikan' (education) and 'bagus' (good) with 28 and 26 occurrences, respectively.

Subsequently, we visualized these eight words in the form of a word cloud. These words collectively represent the community's expression regarding IPDN Jatinangor. The prominence of certain terms, such as 'ipdn', 'tempat', and 'pendidikan', suggests that these concepts are frequently associated with discussions about IPDN Jatinangor among reviews. Furthermore, the presence of positive terms like 'bagus' (good) indicates that there are favorable sentiments expressed alongside constructive feedback and areas for improvement. This wordcloud visualization, as depicted in Fig. 5, provides a concise yet informative representation of the key themes and sentiments conveyed in the reviews, aiding in the comprehensive understanding of public opinion towards IPDN Jatinangor.

Fig. 5. Wordcloud visualization

Meanwhile, the visualization in Fig. 5 offers a comprehensive overview of the most frequently utilized words across all reviews. In this visualization, the size of the letters corresponds to the frequency of usage, with larger letters denoting more frequent occurrence, and progressively smaller letters indicating decreasing frequency. This visual representation enables viewers to quickly grasp the predominant terms used in the reviews, highlighting the key topics and themes discussed by reviewers.

IV. Conclusion

The analysis of text data extracted from Google reviews of IPDN Jatinangor has yielded significant insights into public statements. Utilizing opinion mining techniques, specifically polarization methods, has enabled the discernment of prevailing sentiment tendencies among reviewers, ranging from positive to negative or neutral. These findings offer valuable guidance for IPDN Jatinangor in understanding public perception and devising strategies to enhance service quality and community engagement, thereby fostering a more favorable image.

V. Acknowledgment

We would like to express our utmost appreciation to the Republic of Indonesia Defense University for the invaluable support extended throughout the duration of this research, from its inception to publication. The assistance and resources provided by the university have played a pivotal role in facilitating the successful completion of this study.

References

[1] B. P. Statistik, "Statistik Telekomunikasi Indonesia 2022," pp. 7823–7830, 2023.

[2] J. I. Robbins and C. L. Tieso, "What is Conflict?," *Engag. with Hist. Classr.*, no. October 2013, pp. 27–32, 2021, doi: 10.4324/9781003234906-2.

[3] A. C. Eberendu, "Unstructured Data: an overview of the data of Big Data," *Int. J. Comput. Trends Technol.*, vol. 38, no. 1, pp. 46–50, 2016, doi: 10.14445/22312803/ijctt-v38p109.

[4] C. Blockeel, P. Drakopoulos, N. P. Polyzos, H. Tournaye, and J. A. García-Velasco, "Review the 'peer review,'" *Reprod. Biomed. Online*, vol. 35, no. 6, pp. 747–749, 2017, doi: 10.1016/j.rbmo.2017.08.017.

[5] W. J. Reynolds, "Compliance," *Saf. Heal. Stage*, pp. 66–108, Jan. 2020, doi: 10.4324/9781351136983-4.

[6] Nurnajihah Rosli & Syafiqah Md Nayan, "Why Customer First?," *J. Undergrad. Soc. Sci. Technol.*, vol. 2, no. 2, pp. 505–524, 2020.

[7] A. Agarwal, V. Sharma, G. Sikka, and R. Dhir, “Opinion mining of news headlines using SentiWordNet,” *2016 Symp. Colossal Data Anal. Networking, CDAN 2016*, 2016, doi: 10.1109/CDAN.2016.7570949.

[8] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis,” *Comput. Linguist.*, vol. 35, no. 3, pp. 399–433, 2009, doi: 10.1162/coli.08-012-R1-06-90.

[9] N. A. M. Razali *et al.*, *Opinion mining for national security: techniques, domain applications, challenges and research opportunities*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00536-5.

[10] P. Assiroj, A. Kurnia, and S. Alam, “The performance of Naïve Bayes, support vector machine, and logistic regression on Indonesia immigration sentiment analysis,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 6, pp. 3843–3852, 2023, doi: 10.11591/eei.v12i6.5688.

[11] B. Batrinca and P. C. Treleaven, “Social media analytics: a survey of techniques, tools and platforms,” *AI Soc.*, vol. 30, no. 1, pp. 89–116, 2015, doi: 10.1007/s00146-014-0549-4.

[12] A. Go, R. Bhayani, and L. Huang, “Twitter Sentiment Classification using Distant Supervision,” *Processing*, vol., pp. 1–6, 2009.

[13] F. Romadoni, Y. Umaidah, and B. N. Sari, “Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 2, pp. 247–253, Jul. 2020, doi: 10.32736/sisfokom.v9i2.903.

[14] R. Akbani and N. Japkowicz, “Applying support vector machines to imbalanced datasets,” *Lect. Notes Artif. Intell. (Subseries Lect. Notes Comput. Sci.*, vol. 3201, no. September, pp. 39–50, 2004, doi: 10.1007/978-3-540-30115-8_7.

[15] V. Singh and S. K. Dubey, “Opinion mining and analysis: A literature review,” *Proc. 5th Int. Conf. Conflu. 2014 Next Gener. Inf. Technol. Summit*, pp. 232–239, 2014, doi: 10.1109/CONFLUENCE.2014.6949318.

[16] L. Zhang and B. Liu, *Sentiment Analysis and Opinion Mining*, vol. 3, no. 1959. 2017. doi: 10.1007/978-1-4899-7687-1.

[17] N. Gupta and R. Agrawal, *Application and techniques of opinion mining*. INC, 2020. doi: 10.1016/B978-0-12-818699-2.00001-9.

[18] IPDN, “Institut Pemerintahan Dalam Negeri,” 2024. https://www.ipdn.ac.id/lokasi_ipdn (accessed Jan. 31, 2024).

[19] C. M. Chew, “Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak,” *PLoS One*, vol. 5, no. 11, 2010, doi: 10.1371/journal.pone.0014118.

[20] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, Feb. 2013, doi: 10.1016/j.eswa.2012.07.059.

[21] P. S. Reddy, D. R. Sri, C. S. Reddy, and S. Shaik, “Sentimental Analysis using Logistic Regression,” no. July, 2021, doi: 10.9790/9622-1107023640.

[22] M. Wankhade, A. Chandra, S. Rao, S. Dara, and B. Kaushik, “A Sentiment Analysis of Food Review using Logistic Regression,” vol. 2, no. 7, pp. 251–260, 2017.

[23] M. M. Altawaier and S. Tiun, “Comparison of machine learning approaches on Arabic twitter sentiment analysis,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 1067–1073, 2016, doi: 10.18517/ijaseit.6.6.1456.

[24] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, “Sentiment analysis of Facebook statuses using Naive Bayes Classifier for language learning,” *IISA 2013 - 4th Int. Conf. Information, Intell. Syst. Appl.*, pp. 198–205, 2013, doi: 10.1109/IISA.2013.6623713.

[25] D. A. Muthia, “Sentiment Analysis of Hotel Review Using Naïve Bayes Algorithm and Integration of Information Gain and Genetic Algorithm As Feature Selection Methods,” *Int. Semin. Sci. Issues Trends*, pp. 25–30, 2014.

[26] Martiti and C. Juliane, “Implementation of Naive Bayes Algorithm on Sentiment Analysis Application,” *Proc. 2nd Int. Semin. Sci. Appl. Technol. (ISSAT 2021)*, vol. 207, pp. 193–200, 2021, doi: 10.2991/aer.k.211106.030.

[27] M. Ahmad and S. Aftab, "Analyzing the Performance of SVM for Polarity Detection with Different Datasets," *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 10, pp. 29–36, 2017, doi: 10.5815/ijmecs.2017.10.04.

[28] D. Gunawan and A. Amalia, "The Design of Lexical Database for Indonesian Language," *J. Phys. Conf. Ser.*, vol. 755, no. 1, 2017, doi: 10.1088/1742-6596/755/1/011001.

[29] N. A. Salsabila, Y. A. Winatmoko, A. A. Septiandri, and A. Jamal, "Colloquial Indonesian Lexicon Nikmatun," *2018 Int. Conf. Asian Lang. Process.*, pp. 226–229, 2018.

[30] Gunawan and A. Saputra, "Building synsets for Indonesian WordNet with monolingual lexical resources," *Proc. - 2010 Int. Conf. Asian Lang. Process. IALP 2010*, pp. 297–300, 2010, doi: 10.1109/IALP.2010.69.

[31] F. Bond, L. T. Lim, E. K. Tang, and H. Riza, "The Combined Wordnet Bahasa," *NUSA Linguist. Stud. Lang. around Indones.*, vol. 57, no. December, pp. 83–100, 2014.

[32] O. University, "The Oxford Handbook of Corpus Phonology," *Oxford Handb. Corpus Phonol.*, Aug. 2014, doi: 10.1093/OXFORDHB/9780199571932.001.0001.

[33] J. Dunn, *Natural Language Processing for Corpus Linguistics*. Cambridge University Press, 2022. doi: 10.1017/9781009070447.

[34] F. Koto, "InSet: Indonesia Sentiment Lexicon," 2023. https://github.com/fajri91/InSet (accessed Feb. 01, 2024).

[35] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.

[36] Y. Azhar, "METODE LEXICON-LEARNING BASED UNTUK IDENTIFIKASI TWEET OPINI BERBAHASA INDONESIA," 2017.