Bibliography

[1] Z. Kamble, Q. T. Doan, H. N. Nguyen, Z. Zeng, and L. L. Zihui, "The significance of online reviews for hotel entrepreneurial success," in *Entrepreneurial development and innovation in family businesses and SMEs*, IGI Global, 2020, pp. 75–86.

[2] D. Marthasari and D. C. Widjaja, "Pengaruh ulasan online pada online travel agency (ota) terhadap minat pemesanan kamar hotel di surabaya," *Jurnal Manajemen Perhotelan*, vol. 6, no. 1, pp. 1–8, 2020.

[3] B. Kumar and M. Sadanandam, "A comprehensive review of approaches, methods, and challenges and applications in sentiment analysis," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 14, pp. 6136–6161, 2021.

[4] K. R. Mabokela, T. Celik, and M. Raborife, "Multilingual sentiment analysis for under-resourced languages: A systematic review of the landscape," *IEEE Access*, vol. 11, pp. 15 996–16 020, 2023. DOI: 10.1109/ACCESS.2022.3224136.

[5] K. Shifullah, H. M. Rakibullah, N. Islam, *et al.*, "Classification of hotel reviews using sentiment analysis and machine learning," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, 2022, pp. 710–715. DOI: 10.1109/ICCIT57492.2022.10054884.

[6] W. F. D. Sudiyono, R. Ferdiana, and A. G. Persada, *Implementasi teknik pre-processing untuk analisis sentimen bahasa indonesia berbasis media sosial twitter*, https://etd.repository.ugm.ac.id/penelitian/detail/175543, Skripsi, Program Studi Teknologi Informasi, Universitas Gadjah Mada, 2019.

[7] V. W. D. Thomas and F. Rumaisa, "Analisis sentimen ulasan hotel bahasa indonesia menggunakan support vector machine dan tf-idf," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, p. 1767, 2022.

[8] J. Miharja, J. L. Putra, and N. Hadianto, "Comparison of machine learning classification algorithm on hotel review sentiment analysis (case study: Luminor hotel pecenongan): Komparasi algoritma klasifikasi machine learning pada analisis sentimen review hotel (studi kasus: Luminor hotel pecenongan)," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 59–64, 2020.

[9] A. Hermawan, I. Jowensen, J. Junaedi, *et al.*, "Implementasi text-mining untuk analisis sentimen pada twitter dengan algoritma support vector machine," *JST (Jurnal Sains Dan Teknologi)*, vol. 12, no. 1, pp. 129–137, 2023.

[10] N. N. Hidayati, "Improving aspect-based sentiment analysis for hotel reviews with latent dirichlet allocation and machine learning algorithms," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 9, no. 2, pp. 144–159, 2023.

[11] I. Utama, S. S. Prasetyowati, and Y. Sibaroni, "Multi-aspect sentiment analysis hotel review using rf, svm, and naıve bayes based hybrid classifier," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 630, 2021.

[12] A. K. Qorita and F. Rahma, "Analisis sentimen berdasarkan aspek pada tempat wisata di daerah istimewa yogyakarta menggunakan metode multinomial naıve bayes," *AUTOMATA*, vol. 3, no. 1, 2022.

[13] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," *Information Processing & Management*, vol. 58, no. 5, p. 102 656, 2021.

[14] A. Patel, P. Oza, and S. Agrawal, "Sentiment analysis of customer feedback and reviews for airline services using language representation model," *Procedia Computer Science*, vol. 218, pp. 2459–2467, 2023.

[15] F. A. Bachtiar, W. Paulina, and A. N. Rusydi, "Text mining for aspect based sentiment analysis on customer review: A case study in the hotel industry.," in *IICST*, 2020, pp. 105–112.

[16] Y. A. Singgalen, "Analisis sentimen wisatawan terhadap taman nasional bunaken dan top 10 hotel rekomendasi tripadvisor menggunakan algoritma svm dan dt berbasis crisp-dm," *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 367–379, 2023.

[17] R. Fetra, T. Pradiani, *et al.*, "The influence of price, facilities, and service quality on re-staying interest," *ADI Journal on Recent Innovation*, vol. 4, no. 2, pp. 184–193, 2023.

[18] F. F. Zain and Y. Sibaroni, "Effectiveness of svm method by naã
ve bayes weighting in movie review classification," *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, vol. 5, no. 2, pp. 108–114, 2019.

[19] A. S. Andini, D. T. Murdiansyah, and K. M. Lhaksmana, "Topic classification of islamic question and answer using naıve bayes and tf-idf method," *Computer Engineering and Applications Journal*, vol. 10, no. 3, pp. 151–160, 2021.

[20] B. Harish, K. Kumar, and H. Darshan, "Sentiment analysis on imdb movie reviews using hybrid feature extraction method," 2019.

[21] A. N. Azhar, M. L. Khodra, and A. P. Sutiono, "Multi-label aspect categorization with convolutional neural networks and extreme gradient boosting," in *2019 International Conference on Electrical Engineering and Informatics (ICEEI)*, IEEE, 2019, pp. 35–40.

[22] B. Wilie, K. Vincentio, G. I. Winata, *et al.*, "Indonlu: Benchmark and resources for evaluating indonesian natural language understanding. arxiv," *arXiv preprint arXiv:2009.05387*, 2020.

[23] Z. Ke, J. Sheng, Z. Li, W. Silamu, and Q. Guo, "Knowledge-guided sentiment analysis via learning from natural language explanations," *IEEE Access*, vol. 9, pp. 3570–3578, 2021. DOI: 10.1109/ACCESS.2020.3048088.

[24] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 845–863, 2022. DOI: 10.1109/TAFFC.2020.2970399.

[25] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications," *International Journal of Scientific and Technology Research*, vol. 9, no. 2, pp. 601–609, 2020.

[26] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, 2022.

[27] A. H. Muhendra *et al.*, "Preprocessing of slang words for sentiment analysis on public perceptions in twitter," in *Advances in Sentiment Analysis-Techniques, Applications, and Challenges*, IntechOpen, 2023.

[28] J. Li and A. Engelbrecht, "Advances in sentiment analysis: Techniques, applications, and challenges," 2024.

[29] D. Berrar, "Bayes' theorem and naive bayes classifier," 2019.

[30] P. Pajila, B. Sheena, A. Gayathri, J. Aswini, M. Nalini, and S. S. R, "A comprehensive survey on naive bayes algorithm: Advantages, limitations and applications," in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, 2023, pp. 1228–1234. DOI: 10.1109/ICOSEC58147.2023.10276274.

[31] D. Valkenborg, A.-J. Rousseau, M. Geubbelmans, and T. Burzykowski, "Support vector machines," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 164, no. 5, pp. 754–757, 2023.

[32] Z. Jun, "The development and application of support vector machine," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1748, 2021, p. 052 006.

[33] M. Suyal and P. Goyal, "A review on analysis of k-nearest neighbor classification machine learning algorithms based on supervised learning," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 43–48, 2022.

[34] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 international conference on intelligent computing and control systems (ICCS)*, IEEE, 2019, pp. 1255–1260.

[35] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[36] B. Al Sari, R. Alkhaldi, D. Alsaffar, *et al.*, "Sentiment analysis for cruises in saudi arabia on social media platforms using machine learning algorithms," *Journal of big Data*, vol. 9, no. 1, p. 21, 2022.

[37] B. F. Azevedo, A. M. A. Rocha, and A. I. Pereira, "Hybrid approaches to optimization and machine learning methods: A systematic literature review," *Machine Learning*, vol. 113, no. 7, pp. 4055–4097, 2024.

[38] N. Z. B. Jannah and K. Kusnawi, "Comparison of naıve bayes and svm in sentiment analysis of product reviews on marketplaces," *Sinkron: jurnal dan penelitian teknik informatika*, vol. 8, no. 2, pp. 727–733, 2024.

[39] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[40] S. Rezvani, F. Pourpanah, C. P. Lim, and Q. Wu, "Methods for class-imbalanced learning with support vector machines: A review and an empirical evaluation," *arXiv preprint arXiv:2406.03398*, 2024.

[41] M. Mujahid, E. Kına, F. Rustam, M. Villar, E. Alvarado, I. Diez, *et al.*, "Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. j big data. 2024; 11: 87," *IEEE Access*, vol. 12, pp. 14 050–60, 2024.

[42] V. Putri, M. Masjkur, and C. Suhaeni, "Performance of smote in a random forest and naive bayes classifier for imbalanced hepatitis-b vaccination status," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1863, 2021, p. 012 073.

[43] G. Haixiang *et al.*, "A survey on addressing high-class imbalance in big data," *Journal of Big Data*, 2018. [Online]. Available: https://link.springer.com/article/10.1186/s40537-018-0151-6.

[44] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social network analysis and mining*, vol. 11, no. 1, p. 81, 2021.

[45] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, "Evaluation of sentiment analysis in finance: From lexicons to transformers," *IEEE Access*, vol. 8, pp. 131 662–131 682, 2020. DOI: 10.1109/ACCESS.2020.3009626.

[46] D. E. Cahyani and I. Patasik, "Performance comparison of tf-idf and word2vec models for emotion text classification," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2780–2788, 2021.

[47] K. Gupta, N. Jiwani, and N. Afreen, "A combined approach of sentimental analysis using machine learning techniques," *Revue d'Intelligence Artificielle*, vol. 37, no. 1, p. 1, 2023.

[48] F. T. Saputra, Y. Nurhadryani, S. H. Wijaya, and D. Defina, "Analisis sentimen bahasa indonesia pada twitter menggunakan struktur tree berbasis leksikon," *Jurnal Teknologi Informasi Dan Ilmu Komputer*, vol. 8, no. 1, pp. 135–146, 2021.

[49] H. J. Christanto and Y. A. Singgalen, "Sentiment analysis of customer feedback reviews towards hotel's products and services in labuan bajo," *Journal of Information Systems and Informatics*, vol. 4, no. 4, pp. 805–822, 2022.

[50] IndoNLP Team, *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*, Accessed: July 26, 2025, 2020. [Online]. Available: https://github.com/IndoNLP/indonlu.

[51] IndoNLP Team, *IndoNLU Dataset on Hugging Face*, Accessed: July 26, 2025, 2020. [Online]. Available: https://huggingface.co/datasets/indonlp/indonlu.

[52] A. W. Pradana and M. Hayaty, "The effect of stemming and removal of stopwords on the accuracy of sentiment analysis on indonesian-language texts," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 4, no. 4, pp. 375–380, 2019. DOI: 10.22219/kinetik.v4i4.912. [Online]. Available: http://dx.doi.org/10.22219/kinetik.v4i4.912.

[53] Unknown, "Indonesian words error detection system using nazief–adriani stemmer," in *International Journal of Advanced Computer Science and Applications*, vol. 10, 2019. [Online]. Available: https://thesai.org/Downloads/Volume10No12/Paper_31-Indonesian_Words_Error_Detection_System.pdf.

[54] M. Afrad, D. C. Febrianto, S. Wijayanto, and M. Y. Fathoni, "Sentiment analysis of visitor reviews on baturaden tourist attraction using machine learning methods," *Edu Komputika Journal*, vol. 11, no. 1, pp. 57–64, Aug. 2024. DOI: 10.15294/edukom.v11i1.10561. [Online]. Available: https://journal.unnes.ac.id/journals/edukom/article/view/10561.

[55] F. Panjaitan, W. Ce, *et al.*, "Evaluation of machine learning models for sentiment analysis in the south sumatra governor election using data balancing techniques," *Journal of Information Systems and Informatics*, vol. 7, no. 1, pp. 464–470, 2025. [Online]. Available: https://doi.org/10.51519/journalisi.v7i1.1019.

[56] R. B. Hadiprakoso, H. Setiawan, R. N. Yasa, and Girinoto, "Text preprocessing for optimal accuracy in indonesian sentiment analysis using a deep learning model with word embedding," in *AIP Conference Proceedings*, AIP Publishing LLC, vol. 2680, 2023, p. 020 050.

[57] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

Appendix A: Hotel Review Dataset

Dataset Sample

The following images show a sample view of the Hoasa dataset, which consists of hotel user reviews and is part of the IndoNLU benchmark.

review	ac	air_panas	bau	general	kebersiha	linen	service	sunrise_m	tv	wifi
lumayan nyaman,tp kebersihan kmr mandi perlu ditingkatkan lg biar gk ada kuning2 di sudutnya lbh bgs	neut	neut	neut	neut	neg	neut	neut	neut	neut	neut
kamarnya kurang maintenance. paling baik kamar mandi dibersihkan lebih sering. kamar mandi adalah bagian yang penting untuk pelanggan. kebersihan selimut bed sheet dan handuk juga perlu diperhatikan lebih jauh lagi.	neut	neut	neut	neut	neg	neg	neut	neut	neut	neut
harga terjangkau, kamar luas, ada sofa, semua oke, kecuali kamar mandi. seharian keluar bau tak sedap entah apa masalahnya.	neut	neut	neg	pos	neut	neut	neut	neut	neut	neut
Jaringan WiFi dalam kamar lelet	neut	neut	neut	neut	neut	neut	neut	neut	neut	neg
secara umum bagus, cuma sprei, sarung bantal dan handuk tidak layak...maaf..	neut	neut	neut	pos	neut	neg	neut	neut	neut	neut
Kamar bersih dan nyaman tetapi amenties dari airy tidak disediakan. Begitu juga dengan compliment free snack dari airy	neut	neut	neut	neut	pos	neg	neut	neut	neut	neut