

DAFTAR PUSTAKA

- [1] F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau, and L. Chen, “Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration,” *J. Inf. Technol. Case Appl. Res.*, vol. 25, no. 3, pp. 277–304, 2023, doi: 10.1080/15228053.2023.2233814.
- [2] B. Marr, “A Short History Of ChatGPT: How We Got To Where We Are Today,” *Forbes*. Accessed: Feb. 14, 2025. [Online]. Available: <https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/>
- [3] M. Ilmar Rifaldi, Y. Raymond Ramadhan, and I. Jaelani, “Analisis Sentimen Terhadap Aplikasi Chatgpt Pada Twitter Menggunakan Algoritma Naïve Bayes,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 2, pp. 802–814, 2023.
- [4] A. Salsabila and Muntahanah, “Analisis Sentimen Ulasan Aplikasi ChatGPT Pada Google Play Menggunakan Metode Support Vector Machine,” *Progresif J. Ilm. Komput.*, vol. 20, no. 2, pp. 714–723, 2024, doi: <http://dx.doi.org/10.35889/progresif.v20i2.2074>.
- [5] Y. Akbar, A. N. H. Regita, Sugiyono, and T. Wahyudi, “Analisa Sentimen Pada Media Sosial ‘ X ’ Pencarian Keyword ChatGPT Menggunakan Algoritma K-Nearest Abstrak,” vol. 5, no. 3, pp. 3291–3305, 2024, doi: <https://doi.org/10.35870/jimik.v5i3.1016>.
- [6] N. A. Semary, W. Ahmed, K. Amin, P. Pławiak, and M. Hammad, “Improving sentiment classification using a RoBERTa-based hybrid model,” *Front. Hum. Neurosci.*, vol. 17, no. December, pp. 1–10, 2023, doi: 10.3389/fnhum.2023.1292010.
- [7] H. Patel, D. Singh Rajput, G. Thippa Reddy, C. Iwendi, A. Kashif Bashir, and O. Jo, “A review on classification of imbalanced data for wireless sensor networks,” *Int. J. Distrib. Sens. Networks*, vol. 16, no. 4, 2020, doi: 10.1177/1550147720916404.
- [8] K. Selva, S. Vittayaareekul, and B. Miranda, “Exploring the Efficacy of Meta-Learning: Unveiling Superior Data Diversity Utilization of MAML Over Pre-training,” no. 2022, pp. 1–12, 2025, [Online]. Available: <http://arxiv.org/abs/2501.08506>
- [9] R. A. Levinson, “What are the strategies to handle class imbalance in classification tasks using transformer models?,” Quora. Accessed: Mar. 12, 2025. [Online]. Available: <https://www.quora.com/What-are-the-strategies-to-handle-class->

imbalance-in-classification-tasks-using-transformer-models

- [10] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst. Appl.*, vol. 244, no. November 2023, p. 122778, 2024, doi: 10.1016/j.eswa.2023.122778.
- [11] S. Nikolopoulos, "Sentiment Analysis using Twitter API and RoBERTa model," ODBMS.org. Accessed: Mar. 12, 2025. [Online]. Available: <https://www.odbms.org/2023/02/sentiment-analysis-using-twitter-api-and-roberta-model/>
- [12] GeeksforGeek, "Overview of ROBERTa model," GeeksforGeek. Accessed: Mar. 12, 2023. [Online]. Available: <https://www.geeksforgeeks.org/overview-of-roberta-model/>
- [13] Y. Younes and B. Mathiak, "Handling Class Imbalance when Detecting Dataset Mentions with Pre-trained Language Models," *ICNLSP 2022 - Proc. 5th Int. Conf. Nat. Lang. Speech Process.*, pp. 78–87, 2022.
- [14] S. Sarmuni and A. Tabrani, "Analisis Bibliometrik Penggunaan Chat Gpt Berbasis Data Google Scholar," *J. Sist. Inf. dan Inform.*, vol. 7, no. 2, pp. 164–177, 2024, doi: 10.47080/simika.v7i2.3376.
- [15] T. B. Brown *et al.*, "Language models are few-shot learners -- special version," *Conf. Neural Inf. Process. Syst. (NeurIPS 2020)*, no. NeurIPS, pp. 1–25, 2020.
- [16] M. Javaid, A. Haleem, R. P. Singh, S. Khan, and I. H. Khan, "Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system," *BenchCouncil Trans. Benchmarks, Stand. Eval.*, vol. 3, no. 2, p. 100115, 2023, doi: 10.1016/j.tbench.2023.100115.
- [17] D. Patil, N. L. Rane, and J. Rane, *Emerging and future opportunities with ChatGPT and generative artificial intelligence in various business sectors*, no. 2024.
- [18] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," *FAccT 2021 - Proc. 2021 ACM Conf. Fairness, Accountability, Transpar.*, pp. 610–623, 2021, doi: 10.1145/3442188.3445922.
- [19] K. Paljug, "Social Media: Definition, Importance, Top Websites, and Apps," investopedia. Accessed: Mar. 14, 2025. [Online]. Available: <https://www.investopedia.com/terms/s/social-media.asp>
- [20] J. Douglas, "What are the 6 Types of Social Media?," Truthdigital. Accessed: Mar.

- 14, 2025. [Online]. Available: <https://truthdigital.co.uk/what-are-the-6-types-of-social-media/>
- [21] A. M. A. Ausat, "The Role of Social Media in Shaping Public Opinion and Its Influence on Economic Decisions," *Technol. Soc. Perspect.*, vol. 1, no. 1, pp. 35–44, 2023, doi: 10.61100/tacit.v1i1.37.
- [22] T. Tu, "PUBLIC TRUST IN AI: IMPLICATIONS FOR POLICY AND REGULATION," 2024. [Online]. Available: [https://www.ipsos.com/sites/default/files/ct/news/documents/2024-09/Ipsos Public Trust in AI.pdf](https://www.ipsos.com/sites/default/files/ct/news/documents/2024-09/Ipsos%20Public%20Trust%20in%20AI.pdf)
- [23] N. Ying, "The Evolving Landscape of Web Scraping on Social Media Platforms," UC Berkeley. Accessed: Mar. 19, 2025. [Online]. Available: <https://dlab.berkeley.edu/news/evolving-landscape-web-scraping-social-media-platforms>
- [24] A. Lakshmi, "Ethical Web Scraping: A Practical Guide to Responsible Data Collection," ScraperAPI. Accessed: Mar. 19, 2025. [Online]. Available: <https://www.scraperapi.com/web-scraping/ethical/>
- [25] K. Hewelt, "Understanding big data infrastructure: essentials and challenges," CodiLime. Accessed: Mar. 19, 2025. [Online]. Available: <https://codilime.com/blog/big-data-infrastructure-essentials-and-challenges/>
- [26] S. Aripiyanto, T. Tukino, A. Sufyan, and R. Nandaputra, "Sentimen Analisis Twitter Ibu Kota Negara Nusantara Menggunakan Long Short-Term Memory dan Lexicon Based," *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 12, no. 2, p. 119, 2022, doi: 10.36448/expert.v12i2.2821.
- [27] Y. Yunefri, Y. E. Fadrial, and S. Sutejo, "Chatbot Pada Smart Cooperative Oriented Problem Menggunakan Natural Language Processing dan Naive Bayes Classifier," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2, pp. 131–140, 2021, doi: 10.31539/intecom.v4i2.2704.
- [28] H. E. Rosyadi, F. Amrullah, R. D. Marcus, and R. R. Affandi, "Rancang Bangun Chatbot Informasi Lowongan Pekerjaan Berbasis Whatsapp dengan Metode NLP (Natural Language Processing)," *Briliant J. Ris. dan Konseptual*, vol. 5, no. 3, p. 619, 2020, doi: 10.28926/briliant.v5i3.487.
- [29] G. R. Narayanaswamy, "Exploiting BERT and RoBERTa to Improve Performance for Aspect Based Sentiment Analysis," *Dissertations*, 2021, doi: 10.21427/3w9n-we77.

- [30] G. Srivastava and C. W. Lin, "Guest Editorial Special Section on Fuzzy-Deep Neural Network Learning in Sentiment Analysis," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 1, pp. 1–2, 2025, doi: 10.1109/TFUZZ.2024.3520662.
- [31] S. M. Ferdous, S. N. E. Newaz, S. Bin, and S. Mugdha, "Sentiment Analysis in the Transformative Era of Machine Learning : A Comprehensive Review," vol. 13, no. January, pp. 331–346, 2025, doi: 10.19139/soic-2310-5070-2113.
- [32] C. Brânzilă, "SENTIMENT ANALYSIS AS A TOOL FOR OPTIMIZING EDUCATIONAL STRATEGIES AND BUSINESS COMMUNICATION," no. August 2024, doi: 10.47743/lincu-202.
- [33] L. Ashbaugh and Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers*, vol. 13, no. 12, 2024, doi: 10.3390/computers13120340.
- [34] F. Wang, "Comparative Evaluation of Sentiment Analysis Methods : From Traditional Techniques to Advanced Deep Learning Models," vol. 0, pp. 23–29, 2024, doi: 10.54254/2755-2721/105/2024TJ0056.
- [35] S. Chen, "Sentiment Analysis Techniques for Deep Learning Classification and Comparison," vol. 0, pp. 74–80, 2025, doi: 10.54254/2753-8818/86/2025.20342.
- [36] M. Zhou, "Review of Text Sentiment Analysis Methods," *计算机工程与应用*, vol. 57, no. 12, pp. 11–24, 2021.
- [37] S. Garg, D. Chaturvedi, T. Jain, A. Mishra, and A. Kapoor, "Navigating the Challenges of Sentiment Analysis: Accuracy and Bias," pp. 1–15, 2023, [Online]. Available: <https://doi.org/10.21203/rs.3.rs-2850627/v1>
- [38] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane, "Imbalanced Data Problem in Machine Learning: A Review," *IEEE Access*, vol. 13, no. January, pp. 13686–13699, 2025, doi: 10.1109/ACCESS.2025.3531662.
- [39] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," *J. Biomed. Inform.*, vol. 90, no. November 2018, p. 103089, 2019, doi: 10.1016/j.jbi.2018.12.003.
- [40] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," *Data Democr. Nexus Artif. Intell. Softw. Dev. Knowl. Eng.*, pp. 83–106, 2020, doi: 10.1016/B978-0-12-818366-3.00005-8.
- [41] T. Maier, J. Menold, and C. McComb, "The Relationship Between Performance and Trust in AI in E-Finance," *Front. Artif. Intell.*, vol. 5, no. June, pp. 1–17, 2022, doi:

10.3389/frai.2022.891529.

- [42] S. Ulya, A. Ridwan, W. Cholid Wahyudin, and F. M. Hana, "Text Mining Sentimen Analisis Pengguna Aplikasi Marketplace Tokopedia Berdasar Rating dan Komentar Pada Google Play Store," *J. Bisnis Digit. dan Sist. Inf.*, vol. 3, no. 2, pp. 33–40, 2022, [Online]. Available: <https://ejr.umku.ac.id/index.php/BIDISFO/article/view/1799>
- [43] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [44] Elastic, "What is a web crawler?," Elstic. Accessed: Mar. 12, 2025. [Online]. Available: <https://www.elastic.co/what-is/web-crawler>
- [45] A. F. Nurzaman, "Data Mining – Text Categorizing," Binus University. Accessed: Mar. 19, 2024. [Online]. Available: <https://sis.binus.ac.id/2021/03/26/data-mining-text-categorizing/>
- [46] V. Janev, "Semantic intelligence in big data applications," *Smart Connect. World Technol. Appl. Shap. Futur.*, pp. 71–89, 2021, doi: 10.1007/978-3-030-76387-9_4.
- [47] M. M. Zayed and F. M. Ba-alwi, "Big Data Formation , Reduction , and Its Impact on Sampling : A survey," vol. 3, no. 1, pp. 597–603, 2025.
- [48] O. Nweke, "Integrating Consumer Behavior Tracking , Competitive Analysis , and Smart Integrating Consumer Behavior Tracking , Competitive Analysis , and Algorithms for Smarter Business Strategies," no. March, 2025, doi: 10.7753/IJCATR1404.1001.
- [49] P. Adhikari, P. Hamal, and F. B. Jnr, "The role of big data analytics and management information systems in consumer personalization in U . S . Retail , banking and finance The role of big data analytics and management information systems in consumer personalization in U . S . Retail , banking," no. February, 2025, doi: 10.30574/ijsra.2025.14.2.0388.
- [50] M. Kumar, "Data-Driven Decision Making : How Organizations Use Analytics to Transform their Strategies," no. September 2019, 2025, doi: 10.36948/ijfmr.2019.v01i02.20554.
- [51] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [52] A. Kulkarni and A. Shivananda, *Natural language processing recipes: Unlocking text*

- data with machine learning and deep learning using python*. 2019. doi: 10.1007/978-1-4842-4267-4.
- [53] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM)," *J. GAUSSIAN*, vol. 9, pp. 376–390, 2020.
- [54] D. Varma, A. Nehansh, and P. Swathy, "Data Preprocessing Toolkit : An Approach to Automate Data Preprocessing," *Interantional J. Sci. Res. Eng. Manag.*, vol. 07, no. 03, pp. 1–5, 2023, doi: 10.55041/ijrsrem18270.
- [55] A. Kusuma and H. N. Irmanda, "Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine," pp. 563–574, 2022.
- [56] A. Nurkasanah and M. Hayaty, "Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text," vol. 14, no. 1, pp. 20–27, 2022.
- [57] R. Ramadhan, Y. A. Sari, and P. P. Adikara, "Perbandingan Pembobotan Term Frequency-Inverse Document Frequency dan Term Frequency-Relevance Frequency terhadap Fitur N-Gram pada Analisis Sentimen," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5075–5079, 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10173>
- [58] W. F. Abdillah, A. Premana, and R. M. H. Bhakti, "Analisis Sentimen Penanganan Covid-19 dengan Support Vector Machine: Evaluasi Leksikon dan Metode Ekstraksi Fitur," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 02, pp. 160–170, 2021, doi: 10.46772/intech.v3i02.556.
- [59] Scikit-learn, "train_test_split," Scikit-learn.org. Accessed: Jun. 14, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
- [60] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 423–431, 2022, doi: 10.32672/jnkti.v5i3.4382.
- [61] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [62] Cohen Jacob, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46 ST-A coefficient of agreement for nominal, 1960, [Online]. Available: <http://epm.sagepub.com>

- [63] S. Lee, "Cohen's Kappa Explained: Powerful Tool for Inter-Rater Analysis," Number Analytics. [Online]. Available: <https://www.numberanalytics.com/blog/cohens-kappa-explained-inter-rater-analysis>
- [64] J. Wei and K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 6382–6388, 2019, doi: 10.18653/v1/d19-1670.
- [65] Z. Cai, Z. Li, Y. Liu, L. Guo, and Y. Song, "Ustnlp16 at SemEval-2025 Task 9: Improving Model Performance through Imbalance Handling and Focal Loss," 2025, [Online]. Available: <http://arxiv.org/abs/2505.00021>
- [66] M. Thomson, H. Murfi, and G. Ardaneswari, "BERT-Based Hybrid Deep Learning with Text Augmentation for Sentiment Analysis of Indonesian Hotel Reviews," no. Data, pp. 468–473, 2023, doi: 10.5220/0012127400003541.
- [67] A. Chandrasekar, "Data Augmentation for Multi-task BERT models," 2023.
- [68] S. Y. Feng *et al.*, "A Survey of Data Augmentation Approaches for NLP," *Find. Assoc. Comput. Linguist. ACL-IJCNLP 2021*, pp. 968–988, 2021, doi: 10.18653/v1/2021.findings-acl.84.
- [69] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," *Cognit. Comput.*, vol. 16, no. 1, pp. 45–74, 2024, doi: 10.1007/s12559-023-10179-8.
- [70] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, 2022, doi: 10.1038/s41598-022-09954-8.
- [71] H. Kim, S. Baek, C. Lee, and S. Member, "Artificial Intelligence (AI) Fundamentals for the Display Industry : A Review (August 2024)," *IEEE Open J. Immersive Displays*, vol. 1, no. July, pp. 204–213, 2024, doi: 10.1109/OJID.2024.3470711.
- [72] J. Burden, M. Tešić, L. Pacchiardi, and J. Hernández-Orallo, "Paradigms of AI Evaluation: Mapping Goals, Methodologies and Culture," 2025, [Online]. Available: <http://arxiv.org/abs/2502.15620>
- [73] D. Carty, "Training Data, Validation Data and Test Data in Machine Learning (ML)," APPLAUSE. Accessed: Mar. 12, 2025. [Online]. Available: <https://www.applause.com/blog/training-data-validation-data-vs-test-data/#:~:text=If training and validation,should be unlabeled. Test>
- [74] Eastgate Software, "Understanding Training, Validation, and Testing Data in ML,"

- Eastgate Software. Accessed: Mar. 12, 2025. [Online]. Available: <https://eastgate-software.com/understanding-training-validation-and-testing-data-in-ml/>
- [75] M. K. Wekenborg, S. Gilbert, and J. N. Kather, “Examining human-AI interaction in real- world healthcare beyond the laboratory,” *npj Digit. Med.*, pp. 1–6, 2025, doi: 10.1038/s41746-025-01559-5.
- [76] A. Bhandari, “Understanding & Interpreting Confusion Matrix in Machine Learning (Updated 2024),” *Analytcsvidhya.Com*, no. Updated, p. 1, 2024, [Online]. Available: <https://www.analytcsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/>
- [77] D. M. Powers, “Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation,” *J. Mach. Learn. Technol.*, vol. 2, no. 21, pp. 37–63, 2011, [Online]. Available: <http://www.bioinfo.in/contents.php?id=51>
- [78] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
- [79] A. Vaswani *et al.*, “Attention Is All You Need,” *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manag.*, no. Nips, pp. 4752–4758, Jun. 2017, doi: 10.1145/3583780.3615497.
- [80] Z. A. Guven, “Comparison of BERT Models and Machine Learning Methods for Sentiment Analysis on Turkish Tweets,” *Proc. - 6th Int. Conf. Comput. Sci. Eng. UBMK 2021*, no. October, pp. 98–101, 2021, doi: 10.1109/UBMK52708.2021.9559014.
- [81] D. G. Mandhasiya, H. Murfi, and A. Bustamam, “The hybrid of BERT and deep learning models for Indonesian sentiment analysis,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 1, pp. 591–602, 2024, doi: 10.11591/ijeecs.v33.i1.pp591-602.
- [82] A. Jazuli, Widowati, and R. Kusumaningrum, “Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback,” *Appl. Sci.*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.3390/app15010172.
- [83] H. Murfi, S. Theresia Gowandi, G. Ardaneswari, and S. Nurrohmah, “BERT-based combination of convolutional and recurrent neural network for indonesian sentiment analysis,” *Appl. Soft Comput.*, vol. 151, pp. 1–15, 2024, doi: 10.1016/j.asoc.2023.111112.
- [84] B. Cheang, B. Wei, D. Kogan, H. Qiu, and M. Ahmed, “Language Representation Models for Fine-Grained Sentiment Classification,” 2020, [Online]. Available: <http://arxiv.org/abs/2005.13619>

- [85] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” no. 1, 2019, [Online]. Available: <http://arxiv.org/abs/1907.11692>
- [86] K. Pearce, T. Zhan, A. Komanduri, and J. Zhan, “A Comparative Study of Transformer-Based Language Models on Extractive Question Answering,” 2021, [Online]. Available: <http://arxiv.org/abs/2110.03142>
- [87] K. Pipalia, R. Bhadja, and M. Shukla, “Comparative analysis of different transformer based architectures used in sentiment analysis,” *Proc. 2020 9th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2020*, pp. 411–415, 2020, doi: 10.1109/SMART50582.2020.9337081.
- [88] A. Murarka, B. Radhakrishnan, and S. Ravichandran, “Detection and Classification of mental illnesses on social media using RoBERTa,” 2020, [Online]. Available: <http://arxiv.org/abs/2011.11226>
- [89] M. Beseiso, “Essay scoring tool by employing Roberta architecture,” *ACM Int. Conf. Proceeding Ser.*, pp. 54–57, 2021, doi: 10.1145/3460620.3460630.
- [90] A. Talmor, Y. Elazar, Y. Goldberg, and J. Berant, “Olympics-on what language model pre-training captures,” *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 743–758, 2020, doi: 10.1162/tacl_a_00342.
- [91] I. Staliūnaitė and I. Iacobacci, “Compositional and lexical semantics in RoBERTa, BERT and DistilBERT: A case study on CoQA,” *EMNLP 2020 - 2020 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, no. 7046, pp. 7046–7056, 2020, doi: 10.18653/v1/2020.emnlp-main.573.
- [92] Cardiff NLP, “Twitter-roBERTa-base for Sentiment Analysis - UPDATED (2022),” Hugging Face. Accessed: Jun. 15, 2025. [Online]. Available: <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>
- [93] Papers With Code, “Focal Loss,” Papers With Code. Accessed: Apr. 15, 2025. [Online]. Available: <https://paperswithcode.com/method/focal-loss>
- [94] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, 2020, doi: 10.1109/TPAMI.2018.2858826.
- [95] A. R. Sivakumaran, A. H. Shnain, J. Lande, A. Sindgi, and P. Sruthi, “Sentiment Analysis in Social Media using Bidirectional Long Short Term Memory with Focus Loss function,” in *2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC)*, IEEE, Sep. 2024, pp. 1–5. doi: 10.1109/ICDSCNC62492.2024.10939609.

- [96] Y. Ma, L. Zhao, and J. Hao, "XLP at SemEval-2020 Task 9: Cross-lingual Models with Focal Loss for Sentiment Analysis of Code-Mixing Language," *14th Int. Work. Semant. Eval. SemEval 2020 - co-located 28th Int. Conf. Comput. Linguist. COLING 2020, Proc.*, pp. 975–980, 2020, doi: 10.18653/v1/2020.semeval-1.126.
- [97] L. He, "Enhanced twitter sentiment analysis with dual joint classifier integrating RoBERTa and BERT architectures," *Front. Phys.*, vol. 12, no. December, pp. 1–12, 2024, doi: 10.3389/fphy.2024.1477714.
- [98] P. Xue and W. Bai, "A Fine-Grained Sentiment Analysis Method Using Transformer for Weibo Comment Text," *Int. J. Inf. Technol. Syst. Approach*, vol. 17, no. 1, pp. 1–24, 2024, doi: 10.4018/ijitsa.345397.
- [99] A. Zela, T. Elsken, T. Saikia, Y. Marrakchi, T. Brox, and F. Hutter, "Understanding and Robustifying Differentiable Architecture Search," *8th Int. Conf. Learn. Represent. ICLR 2020*, no. Section 5, 2020.
- [100] M. Claesen and B. De Moor, "Hyperparameter Search in Machine Learning," pp. 10–14, 2015, [Online]. Available: <http://arxiv.org/abs/1502.02127>
- [101] Mathieu, "The History of Python Programming," BOCASAY. [Online]. Available: <https://www.bocasay.com/history-python-programming/#:~:text=Python was created by,a hobby project to>
- [102] M. Saeed, "More Special Features in Python," Machine Learning Mastery. [Online]. Available: <https://machinelearningmastery.com/python-special-features/#:~:text=Python has awesome special,zip%2C enumerate%2C decorators and>
- [103] A. Lheureux, "Introduction to Natural Language Processing using NLTK," DigitalOcean. [Online]. Available: <https://blog.paperspace.com/introduction-to-natural-language-processing-using-nltk/#:~:text=semantic reasoning%2C it offers,lexical database of semantic>
- [104] GeeksforGeeks, "Introduction to NLTK: Tokenization, Stemming, Lemmatization, POS Tagging," GeeksforGeeks. Accessed: Apr. 12, 2025. [Online]. Available: <https://www.geeksforgeeks.org/introduction-to-nltk-tokenization-stemming-lemmatization-pos-tagging/>
- [105] GeeksforGeeks, "NLTK Sentiment Analysis Tutorial For Beginners," GeeksforGeeks. Accessed: Apr. 12, 2025. [Online]. Available: <https://www.geeksforgeeks.org/nltk-sentiment-analysis-tutorial-for-beginners/>
- [106] GeeksforGeeks, "What is python scikit library?," GeeksforGeeks. Accessed: Apr. 12,

2025. [Online]. Available: <https://www.geeksforgeeks.org/what-is-python-scikit-library/>
- [107] Scikit-learn, “TfidfVectorizer,” Scikit-learn.org. Accessed: Apr. 14, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
- [108] Hardikkumar, “Performing Sentiment Analysis With Naive Bayes Classifier!,” Analytics Vidhya. Accessed: Apr. 14, 2025. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/07/performing-sentiment-analysis-with-naive-bayes-classifier/#:~:text=Naive Bayes is the,analysis%2C and recommendation systems%2C>
- [109] D. Gera and A. Kapoor, “Sentiment Analysis using Scikit Learn: A Review,” *Researchgate.Net*, no. February, 2022, doi: 10.13140/RG.2.2.26189.10720.
- [110] W. Nugraha and R. Sabaruddin, “Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM Resampling Technique for Handling Class Imbalance in the Classification of Diabetes using C4.5, Random Forest, and SVM,” *Agustus*, vol. 20, no. 3, pp. 352–361, 2021, [Online]. Available: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>.
- [111] M. McDougall, “Text Data Augmentation Techniques,” Mervin McDougall. Accessed: Aug. 08, 2025. [Online]. Available: <https://www.mervinmcdougall.com/blog/2025-05/text-data-augmentation-techniques>