DAFTAR PUSTAKA

Agarap, A. F. M. (2019). *On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset*. arxiv. https://doi.org/10.48550/arXiv.1711.07831

Aljabri dkk. (2023). Machine learning-based social media bot detection: a comprehensive literature review. *Social Network Analysis and Mining, 13*(1). https://doi.org/10.1007/s13278-022-01020-5

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*, 53. https://doi.org/10.1186/s40537-021-00444-8

Amorim, L. B. V., Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). The choice of scaling technique matters for classification performance. *Applied Soft Computing, 133*, 109924. https://doi.org/10.1016/j.asoc.2022.109924

Antonakaki, D., Fragopoulou, P., & Ioannidis, S. (2021). A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications, 164*, 114006. https://doi.org/10.1016/j.eswa.2020.114006

Associated Press. (2022). *Elon Musk says Twitter deal on hold over fake account data*. CBS News. https://www.cbsnews.com/news/elon-musk-twitter-bots-new-owner-2022/. Diakses pada 21 Februari 2026.

Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review*, *56,* 10345–10425. https://doi.org/10.1007/s10462-023-10419-1

Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Benabbou, F., Boukhouima, H., & Sael, N. (2022). Fake accounts detection system based on bidirectional gated recurrent unit neural network. *International Journal of Electrical and Computer Engineering, 12*(3), 3129–3137. https://doi.org/10.11591/ijece.v12i3.pp3129-3137

Buigues, P.J., Arnau, V., Fernández-de-Córdoba, P., & Iglesias, A. I. (2020). Effect of sequence padding on the performance of deep learning models in archaeal protein functional prediction. *Scientific Reports, 10,* 14634. https://doi.org/10.1038/s41598-020-71450-8

Bishop, C. M. (2006). Pattern Recognition and *Machine learning*. Springer.

Chatterjee, A., Saha, J., & Mukherjee, J. (2022). Clustering with multi-layered perceptron. *Pattern Recognition Letters,* 155, 92–99.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing, 9*(6), 811–824.

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. *Proceedings of the 26th International World Wide Web Conference Companion (WWW '17 Companion)*, 963–972. https://doi.org/10.1145/3041021.3055135

Cresci, S. (2020). A Decade of Social Bot Detection. *Communications of the ACM, 63*(10), 72–83. https://doi.org/10.1145/3409116

Cover, T. M., & Thomas, J. A. (2006). Elements of information theory (2nd ed.). *Hoboken, NJ: Wiley-Interscience*.

DataReportal. (2025). *Essential X (Twitter) statistics and trends*. https://datareportal.com/essential-twitter-stats. Diakses pada 21 Februari 2026.

Daouadi, K. E., Zghal Rebaï, R., & Amous, I. (2019). Bot detection on online social networks using deep forest. In R. Silhavy (Ed*.), Computational Science and Its Applications (AISC 985)* (pp. 307–315). Springer. https://doi.org/10.1007/978-3-030-19810-7_30

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Márquez, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., & Reineking, B. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2021). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing, 503,* 92–108.

Ellaky, Z., Benabbou, F., & Ouahabi, S. (2023). Systematic literature review of social media bots detection systems. *Journal of King Saud University – Computer and Information Sciences, 35,* 101551. https://doi.org/10.1016/j.jksuci.2023.04.004

Ellaky, Z., Benabbou, F., Matrane, Y., & Qaqa, S. (2024). A hybrid deep learning architecture for social media bots detection based on BiGRU-LSTM and GloVe word embedding. *IEEE Access, 12*, 100278-100294. https://doi.org/10.1109/ACCESS.2024.3430859

Fachrurrazi. (2025). New public sphere di Indonesia: Peran media sosial dalam mengkonstruksi diskursus publik di dunia maya. *Jurnal Review Pendidikan dan Pengajaran, 8*(1), 804–810.

Feng, S., Wan, H., Wang, N., Li, J., & Luo, M. (2021). TwiBot-20: A comprehensive Twitter bot detection benchmark. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21)* (pp. 448–457). Association for Computing Machinery. https://doi.org/10.1145/3459637.3482019

Feng, S., Tan, Z., Wan, H., Wang, N., Chen, Z., Zhang, B., Zheng, Q., Zhang, W., Lei, Z., Yang, S., dkk. (2022). TwiBot-22: Towards graph-based Twitter bot detection. *Advances in Neural Information Processing Systems, 35*, 35254–35269. https://doi.org/10.48550/arXiv.2206.04564

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research, 61*, 863–905. https://doi.org/10.1613/jair.1.11192

Ferrara, E. (2018). Measuring social spam and the effect of bots on information diffusion in social media. In *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks* (pp. 229–255). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-77332-2_13

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine learning Research, 3*, 1289–1305.

Gali, N., dkk. (2024). Twitter spam detection using deep learning and Word2Vec. *Journal of Information and Communication Technology*, *12*(3), 4213–4220.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: Methods and prospects. *Big Data Analytics, 1*, 9. https://doi.org/10.1186/s41044-016-0014-0

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). *O'Reilly Media*.

Goldberg, Y. (2017). *Neural network methods for natural language processing*. Morgan & Claypool Publishers. https://doi.org/10.2200/S00762ED1V01Y201703HLT037

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. *MIT Press*.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). Survey on categorical data for neural networks. *Journal of Big Data, 7*, 28. https://doi.org/10.1186/s40537-020-00305-w

Hasanin, T., Khoshgoftaar, T. M., Leevy, J. L., & Bauder, R. A. (2019). Severely imbalanced big data challenges: Investigating data sampling approaches. *Journal of Big Data, 6,* 107. https://doi.org/10.1186/s40537-019-0274-4

Hayawi, K., Kaosar, M., Saha, S., Masud, M. M., & Mathew, S. S. (2023). Social media bot detection with deep learning methods: A systematic review. *Neural Computing and Applications, 35,* 12843–12872.

Heidari, M., & Jones, J. H. Jr. (2020). Using BERT to extract topic-independent sentiment features for social media bot detection. *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 542-547). https://doi.org/10.1109/UEMCON51285.2020.9298158

Hickey, D., Fessler, D. M. T., Lerman, K., & Burghardt, K. (2025). X under Musk's leadership: Substantial hate and no reduction in inauthentic activity. *PLOS ONE, 20*(2), e0313293. https://doi.org/10.1371/journal.pone.0313293

Hollenbaugh, E. E. (2021). Self-presentation in social media: Review and research opportunities. *Review of Communication Research, 9,* 80–98.

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *npj Digital Medicine, 3*, 136. https://doi.org/10.1038/s41746-020-00341-z

Ilias, L., Kazelidis, I. M., & Askounis, D. (2024). Multimodal detection of bots on X (Twitter) using transformers. *IEEE Transactions on Information Forensics and Security*, *19*, 7320–7334, 2024, doi: 10.1109/TIFS.2024.3435138.

Ilias, L., & Roussaki, I. (2021). Detecting malicious activity in Twitter using deep learning techniques. *Applied Soft Computing, 107,* 107360.

Kaya, M. O. (2021). Classification of heart failure patients using multi-layer perceptron. *Biomedical Signal Processing and Control*, *6*(1), 35–38. https://doi.org/10.52876/jcs.913671.

Koukaras, P., & Tjortjis, C. (2025). Data preprocessing and feature engineering for data mining: Techniques, tools, and best practices. *AI*, *6*(1), 257. https://doi.org/10.3390/ai6100257

Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. *Information Sciences, 467,* 312–322.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. New York, NY: Springer. https://doi.org/10.1007/978-1-4614-6849-3

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine learning Research, 18*(185), 1–52.

Li, Y., Yang, Y., Song, P., Duan, L., & Ren, R. (2025). An improved SMOTE algorithm for enhanced imbalanced data classification by expanding sample generation space. *Scientific Reports, 15,* 1–15. https://doi.org/10.1038/s41598-025-095

Lin, S.-Y., Chen, Y.-C., Chang, Y.-H., Lo, S.-H., & Chao, K.-M. (2024). Text–image multimodal fusion model for enhanced fake news detection. *Science Progress, 107*(4). https://doi.org/10.1177/00368504241292685

Liu, Y., Kliman-Silver, C., & Mislove, A. (2021). The tweets they are a-changin': Evolution of Twitter users and behavior. *Proceedings of the International AAAI Conference on Web and Social Media, 15,* 305–314.

Liu, Y., & Zhao, H. (2017). Variable importance-weighted random forests. *Quantitative Biology, 5*(4), 338–351.

Lopez-del Rio, A., Martin, M., Perera-Lluna, A., Saidi, R. (2020). Effect of sequence padding on the performance of deep learning models in archaeal protein functional prediction. *Scientific Reports, 10,* 14634. https://doi.org/10.1038/s41598-020-71450-8

Manning, C. D., Raghavan, P., & Schütze, H. (2009). An introduction to information retrieval. *Cambridge University Press*.

Mazza, M., Cola, G., & Tesconi, M. (2022). Ready-to-(ab)use: From fake account trafficking to coordinated inauthentic behavior on Twitter. *Online Social Networks and Media, 31*, 100224. https://doi.org/10.1016/j.osnem.2022.100224

Mujahid, M., Kına, E., & Rustam, F. (2024). Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. *Journal of Big Data, 11,* 87.

Mumuni, A., & Mumuni, F. (2024). Automated data processing and feature engineering for deep learning and big data applications: A survey. *Artificial Intelligence Review, 56,* 10429–10471. https://doi.org/10.1007/s10462-023-10423-5

Mohanapriya, K., Sangavi, N., Kanimozhi, A., Kiruthika, V. R., & Dhivya, P. (2023). Optimized feed forward neural network for fake and clone account detection in online social networks. In *Proceedings of ICSCDS 2023.* IEEE.

Mondal, H., Atanasov, A. G., Eibensteiner, F., Hribersek, M., Brandstätter, S., Matin, M., dkk. (2023). Science communication on X (formerly Twitter): A picture is worth a thousand

characters? *Exploration of Digital Health Technologies, 1*, 28–34. https://doi.org/10.37349/edht.2023.00005

Pal, U., Mondal, S., & Mondal, M. A. (2025). A hybrid CNN–LSTM approach for accident tweet classification. *International Research Journal of Multidisciplinary Scope, 6*(3), 1153–1167. https://doi.org/10.47857/irjms.2025.v06i03.04896

Pattanayak, S., & Swamynathan, M. (2019). *Intelligent projects using Python*. Packt Publishing.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 1532–1543.

Priya C., B. (2025). *MinMax vs Standard vs Robust Scaler: Which One Wins for Skewed Data?* Machine Learning Mastery. Diakses pada 5 Februari 2026.

Qixuan, Y. (2024). *Three-class text sentiment analysis based on LSTM*. arXiv. https://arxiv.org/abs/2412.17347

Rafat, Y., Narayana, P., Mohana, R.M., & Srilatha, K. (2025). LSTM-Based News Article Category Classification. *Computer Science & Mathemathics Forum, 12,* 8. https://doi.org/10.3390/cmsf2025012008

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports, 14,* 6086. https://doi.org/10.1038/s41598-024-56706-x

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2*, 160.

Sengar, S. S., Kumar, S., Raina, P., & Mahaliyan, M. (2020). Bot detection in social networks based on multilayered deep learning approach. *Sensors & Transducers, 244*(5), 37–43.

Sharma, S., & Gupta, V. (2022). Role of Twitter user profile features in retweet prediction for big data streams. *Multimedia Tools and Applications, 81*, 27309–27338. https://doi.org/10.1007/s11042-022-12815-1

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2018). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology, 10*(3), Article 21.

Shen, X., Huang, M., Hu, Z., Cai, S., & Zhou, T. (2024). Multimodal fake news detection with contrastive learning and optimal transport. *Frontiers in Computer Science, 6*, 1473457. https://doi.org/10.3389/fcomp.2024.1473457

Snoek, C. G. M., Worring, M., & Smeulders, A. W. M. (2005). Early versus late fusion in semantic video analysis. *Proceedings of the 13th Annual ACM International Conference on Multimedia (MM '05)*. https://doi.org/10.1145/1101149.1101236

Terumalasetti, S., & Reeja, S. R. (2024). Enhancing social media user's trust: A comprehensive framework for detecting malicious profiles using multi-dimensional analytics. *IEEE Access, 13,* 7071–7093.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. *Proceedings of ICWSM*.

Wei, F., & Nguyen, U. T. T. (2019). Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings. *IEEE TPS-ISA 2019.*

Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'Ban, S. R., Amalia, M., & Febriansyah, R. (2024). Pembelajaran machine learning. *OKTAL: Jurnal Ilmu Komputer dan Science, 3*(2), 375–380.

Wu, B., Liu, L., Yang, Y., Zheng, K., & Wang, X. (2020). Using improved conditional generative adversarial networks to detect social bots on Twitter. *IEEE Access, 8,* 36664–36676.

Yang, C., Harkreader, R., & Gu, G. (2013). Empirical evaluation and new design for fighting evolving Twitter spammers. *IEEE Transactions on Information Forensics and Security, 8*(8), 1280–1293.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys, 51*(2), 32. https://doi.org/10.1145/3161603