DAFTAR PUSTAKA

Alkaberi, W. S., Aljuhani, R. H., & Alamoudi, H. M. (2021). Web Scraper Application for Extracting Scientific Journals Data. *The 5th International Conference on Future Networks & Distributed Systems*, 220–224. https://doi.org/10.1145/3508072.3508106

Anandarajan, M., Hill, C., & Nolan, T. (2019). *Text Preprocessing* (hlm. 45–59). https://doi.org/10.1007/978-3-319-95663-3_4

Bandung, K. (2021). *KONDISI GEOGRAFI KOTA BANDUNG*.

Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies (ICOT)*, 1–6. https://doi.org/10.1109/ICOT.2018.8705796

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *14*(1), 106. https://doi.org/10.1186/1471-2105-14-106

Chaerul, M., Fahruroji, A. R., & Fujiwara, T. (2014). Recycling of plastic packaging waste in Bandung City, Indonesia. *Journal of Material Cycles and Waste Management*, *16*(3), 509–518. https://doi.org/10.1007/s10163-013-0201-2

Chauhan, N., Singh, M., Verma, A., Parasher, A., & Budhiraja, G. (2019). Implementation of database using python flask framework. *International Journal of Engineering and Computer Science*, *8*(12), 24894–24899. https://doi.org/10.18535/ijecs/v8i12.4390

Data Miner. (2023). https://dataminer.io/

de Andrade, B. M., de Gois, J. S., Xavier, V. L., & Luna, A. S. (2020). Comparison of the performance of multiclass classifiers in chemical data: Addressing the problem of overfitting with the permutation test. *Chemometrics and Intelligent Laboratory Systems*, *201*, 104013. https://doi.org/10.1016/j.chemolab.2020.104013

Devndra Ghimire. (2020). *Comparative study on Python web frameworks: Flask and Django*. Metropolia University of Applied Sciences.

Dewan Perwakilan Rakyat Republik Indonesia. (2022). *Ditjen PSLB3 KLHK Didesak Miliki Langkah Terukur Tangani Volume Sampah*. https://www.dpr.go.id/berita/detail/id/40924/t/Ditjen+PSLB3+KLHK+Dide sak+Miliki+Langkah+Terukur+Tangani+Volume+Sampah

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019). Web Scraping: State-of-the-Art and Areas of Application. *2019 IEEE International Conference on Big Data (Big Data)*, 6040–6042. https://doi.org/10.1109/BigData47090.2019.9005594

Duong, H.-T., & Nguyen-Thi, T.-A. (2021). A review: preprocessing techniques and data augmentation for sentiment analysis. *Computational Social Networks*, *8*(1), 1. https://doi.org/10.1186/s40649-020-00080-x

Elreedy, D., & Atiya, A. F. (2019). *A Novel Distribution Analysis for SMOTE Oversampling Method in Handling Class Imbalance* (hlm. 236–248). https://doi.org/10.1007/978-3-030-22744-9_18

Elyan, E., Moreno-Garcia, C. F., & Jayne, C. (2021). CDSMOTE: class decomposition and synthetic minority class oversampling technique for imbalanced-data classification. *Neural Computing and Applications*, *33*(7), 2839–2851. https://doi.org/10.1007/s00521-020-05130-z

Ertam, F. (2018). Deep learning based text classification with Web Scraping methods. *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1–4. https://doi.org/10.1109/IDAP.2018.8620790

Ferronato, N., & Torretta, V. (2019). Waste Mismanagement in Developing Countries: A Review of Global Issues. *International Journal of Environmental Research and Public Health*, *16*(6), 1060. https://doi.org/10.3390/ijerph16061060

Figueiredo, L. N. L., de Assis, G. T., & Ferreira, A. A. (2017). DERIN: A data extraction method based on rendering information and n-gram. *Information Processing & Management*, *53*(5), 1120–1138. https://doi.org/10.1016/j.ipm.2017.04.007

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm. *Procedia Computer Science*, *161*, 765–772. https://doi.org/10.1016/j.procs.2019.11.181

García, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining* (Vol. 72). Springer International Publishing. https://doi.org/10.1007/978-3-319-10247-4

Geoffrey I Webb. (2017). *Encyclopedia of Machine Learning and Data Mining*. Springer US. https://doi.org/10.1007/978-1-4899-7687-1

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. Dalam *ACM Computing Surveys* (Vol. 49, Nomor 2). Association for Computing Machinery. https://doi.org/10.1145/2938640

Gosain, A., & Sardana, S. (2017). Handling class imbalance problem using oversampling techniques: A review. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 79–85. https://doi.org/10.1109/ICACCI.2017.8125820

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, *117*, 47–58. https://doi.org/10.1016/j.future.2020.11.022

Han, J. (2000). *Data Mining Third Edition*.

Hotz, N. (2023, Januari 19). *What is CRISP DM? - Data Science Process Alliance*. https://www.datascience-pm.com/crisp-dm-2/

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, *79*, 403–408. https://doi.org/10.1016/j.procir.2019.02.106

Hudaya, C. S., Fakhrurroja, H., & Alamsyah, A. (2019). ANALISIS PERSEPSI KONSUMEN TERHADAP BRAND GO-JEK PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE SENTIMENT ANALYSIS DAN TOPIC MODELLING. *Jurnal Mitra Manajemen*, *3*(6), 664–673. https://doi.org/10.52160/ejmm.v3i6.244

IBM. (2023). IBM SPSS Modeler CRISP-DM Guide. Dalam *IBM SPSS Modeler CRISP-DM Guide*.

Jo, T. (2019). *Text Mining* (Vol. 45). Springer International Publishing. https://doi.org/10.1007/978-3-319-91815-0

Joval Ifghaniyafi Farras, Sarasi, V., Chaerudin, I., Primiana, I., & Yunani, A. (2022). Implementation of lean process to solid waste management in Bandung, Indonesia. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management)*, *12*(2), 210–227. https://doi.org/10.29244/jpsl.12.2.210-227

JustAnotherArchivist. (2023). *snscrape: A social networking service scraper in Python*. https://github.com/JustAnotherArchivist/snscrape

Kabirifar, K., Mojtahedi, M., Wang, C., & Tam, V. W. Y. (2020). Construction and demolition waste management contributing factors coupled with reduce, reuse, and recycle strategies for effective waste management: A review. *Journal of Cleaner Production*, *263*, 121265. https://doi.org/10.1016/j.jclepro.2020.121265

Kandhro, I. A., Jumani, S. Z., Lashari, A. A., Nangraj, S. S., Lakhan, Q. A., Baig, M. T., & Guriro, S. (2019). Classification of Sindhi Headline News Documents based on TF-IDF Text Analysis Scheme. *Indian Journal of Science and Technology*, *12*(33), 1–10. https://doi.org/10.17485/ijst/2019/v12i33/146130

SIPSN. (2022). *SIPSN - Sistem Informasi Pengelolaan Sampah Nasional*. https://sipsn.menlhk.go.id/sipsn/public/data/capaian

Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia. (2022). *Status Lingkungan Hidup Indonesia 2022*.

Klosterman, S. (2021). *Data Science Projects with Python Second Edition* (S. Klosterman, Ed.; 2 ed.). Packt Publishing Ltd.

Kotha, U. M., Gaddam, H., Siddenki, D. R., & Saleti, S. (2023). A comparison of various machine learning algorithms and execution of flask deployment on

essay grading. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(3), 2990. https://doi.org/10.11591/ijece.v13i3.pp2990-2998

Kramer, O. (2016). *Scikit-Learn* (hlm. 45–53). https://doi.org/10.1007/978-3-319-33383-0_5

Krawczyk, B., & Wozniak, M. (2019). *On the Role of Cost-Sensitive Learning in Imbalanced Data Oversampling* (hlm. 180–191). https://doi.org/10.1007/978-3-030-22744-9_14

ksnugroho. (2019). *klasifikasi-spam-sms*. GitHub. https://github.com/ksnugroho/klasifikasi-spam-sms/tree/master/data

Kurniawan, T. A., Meidiana, C., Dzarfan Othman, M. H., Goh, H. H., & Chew, K. W. (2023). Strengthening waste recycling industry in Malang (Indonesia): Lessons from waste management in the era of Industry 4.0. *Journal of Cleaner Production*, *382*, 135296. https://doi.org/10.1016/j.jclepro.2022.135296

Lagman, A. C., Calleja, J. Q., Fernando, C. G., Gonzales, J. G., Legaspi, J. B., Ortega, J. H. J. C., Ramos, R. F., Solomo, M. V. S., & Santos, R. C. (2019). Embedding naïve Bayes algorithm data model in predicting student graduation. *Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering*, 51–56. https://doi.org/10.1145/3369555.3369570

Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes. *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, 49–54. https://doi.org/10.1109/ICTS.2019.8850982

Mahmoud, A. B., Hack-Polay, D., Grigoriou, N., Mohr, I., & Fuxman, L. (2021). A generational investigation and sentiment and emotion analyses of female fashion brand users on Instagram in Sub-Saharan Africa. *Journal of Brand Management*, *28*(5), 526–544. https://doi.org/10.1057/s41262-021-00244-8

Marešová, D., Mareš, K., Alexiou-Ivanova, T., Satyakti, Y., & Pilařová, T. (2023). Evaluating the waste management situation and attitudes of residents in Bandung, Indonesia. *Integrated Environmental Assessment and Management*, *19*(1), 114–125. https://doi.org/10.1002/ieam.4630

McCabe, S. D., Lin, D.-Y., & Love, M. I. (2020). Consistency and overfitting of multi-omics methods on experimental data. *Briefings in Bioinformatics*, *21*(4), 1277–1284. https://doi.org/10.1093/bib/bbz070

Mohamed Ridhwan, K., & Hargreaves, C. A. (2021). Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *International Journal of Information Management Data Insights*, *1*(2), 100021. https://doi.org/10.1016/j.jjimei.2021.100021

Nandal, N., Tanwar, R., & Pathan, A.-S. K. (2023). Sentiment Analysis based Emotion Extraction for COVID-19 Using Crawled Tweets and Global Statistics for Mental Health. *Procedia Computer Science*, *218*, 949–958. https://doi.org/10.1016/j.procs.2023.01.075

Nasteski, V. (2017). An overview of the supervised machine learning methods. *HORIZONS.B*, *4*, 51–62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05

PANG-NING TAN, MICHAEL STEINBACH, & VIPIN KUMAR. (2006). *INTRODUCTION TO DATA MINING* (M. Goldstein, Ed.). Pearson Education, Inc.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12*.

Pemerintah Indonesia. (2008). *UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 18 TAHUN 2008 TENTANG PENGELOLAAN SAMPAH*.

Pillai, P., & Amin, D. (2020). Understanding the requirements Of the Indian IT industry using web scrapping. *Procedia Computer Science*, *172*, 308–313. https://doi.org/10.1016/j.procs.2020.05.050

Pokhrel, S., Abbas, R., & Aryal, B. (2021). *IoT Security: Botnet detection in IoT using Machine learning*.

Pradha, S., Halgamuge, M. N., & Tran Quoc Vinh, N. (2019). Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data. *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, 1–8. https://doi.org/10.1109/KSE.2019.8919368

Rachmawati, E., Alamianti, D., Kamil, R. I., Suhara, E., & Yani, W. O. N. (2019). The implementation of local government's policy using the waste management program based on household needs in Bandung city. *International Journal of Engineering and Advanced Technology*, *8*(5), 521–528. https://doi.org/10.35940/ijeat.E1075.0585C19

Raharjo, B. (2017). *Belajar Otodidak Flask Framework Python untuk Pengembangan Aplikasi Web* (B. Raharjo, Ed.; 1 ed.). Informatika Bandung.

Raharjo, B. (2019). *Mudah Belajar Python Untuk Aplikasi Desktop dan Web* (B. Raharjo, Ed.; Revisi). Informatika Bandung.

Ramesh Sharda, Dursun Delen, & Efraim Turban. (2016). *BUSINESS INTELLIGENCE, ANALYTICS, AND DATA SCIENCE: A Managerial Perspective* (4 ed.). Pearson Education, Inc.

Ramos, J. (2003). Using TF-IDF to Determine Word Relevance in Document Queries . *Proceedings of the first instructional conference on machine learning*, *242*(1), 29–48.

Rosid, M. A., Fitrani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, *874*(1), 012017. https://doi.org/10.1088/1757-899X/874/1/012017

Saad, S. E., & Yang, J. (2019). Twitter Sentiment Analysis Based on Ordinal Regression. *IEEE Access*, *7*, 163677–163685. https://doi.org/10.1109/ACCESS.2019.2952127

Sabrani, A., Wedashwara W., I. G. W., & Bimantoro, F. (2020). Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia. *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTIKA)*, *2*(1), 89–100. https://doi.org/10.29303/jtika.v2i1.87

Samaras, L., García-Barriocanal, E., & Sicilia, M.-A. (2023). Sentiment analysis of COVID-19 cases in Greece using Twitter data. *Expert Systems with Applications*, *230*, 120577. https://doi.org/10.1016/j.eswa.2023.120577

Siddiqua, A., Hahladakis, J. N., & Al-Attiya, W. A. K. A. (2022). An overview of the environmental pollution and health effects associated with waste landfilling and open dumping. *Environmental Science and Pollution Research*, *29*(39), 58514–58536. https://doi.org/10.1007/s11356-022-21578-z

Solano, J. A., Lancheros Cuesta, D. J., Umaña Ibáñez, S. F., & Coronado-Hernández, J. R. (2022). Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test. *Procedia Computer Science*, *198*, 512–517. https://doi.org/10.1016/j.procs.2021.12.278

Sontakke, S. A., Lohokare, J., Dani, R., & Shivagaje, P. (2019). *Classification of Cardiotocography Signals Using Machine Learning* (hlm. 439–450). https://doi.org/10.1007/978-3-030-01057-7_35

Soufyane, A., Abdelhakim, B. A., & Ahmed, M. Ben. (2021). *An Intelligent Chatbot Using NLP and TF-IDF Algorithm for Text Understanding Applied to the Medical Field* (hlm. 3–10). https://doi.org/10.1007/978-3-030-53440-0_1

Thomas, D. M., & Mathur, S. (2019). Data Analysis by Web Scraping using Python. *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, 450–454. https://doi.org/10.1109/ICECA.2019.8822022

Tonkin, E. L. (2016). Working with Text. Dalam *Working with Text* (hlm. 1–21). Elsevier. https://doi.org/10.1016/B978-1-84334-749-1.00001-9

Unpingco, J. (2021). *Python Programming for Data Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-030-68952-0

Uther, W., Mladenić, D., Ciaramita, M., Berendt, B., Kołcz, A., Grobelnik, M., Mladenić, D., Witbrock, M., Risch, J., Bohn, S., Poteet, S., Kao, A., Quach,

L., Wu, J., Keogh, E., Miikkulainen, R., Flener, P., Schmid, U., Zheng, F., … Nijssen, S. (2011). TF–IDF. Dalam *Encyclopedia of Machine Learning* (hlm. 986–987). Springer US. https://doi.org/10.1007/978-0-387-30164-8_832

Wahid, D. H., & SN, A. (2016). Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *10*(2), 207. https://doi.org/10.22146/ijccs.16625

Wang, H., He, J., Zhang, X., & Liu, S. (2020). A Short Text Classification Method Based on N-Gram and CNN. *Chinese Journal of Electronics*, *29*(2), 248–254. https://doi.org/10.1049/cje.2020.01.001

We Are Social. (2023). *The Changing World of Digital In 2023*. https://wearesocial.com/us/blog/2023/01/the-changing-world-of-digital-in-2023/

Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, *44*(1), 48–59. https://doi.org/10.1177/0165551516677946

Zhang, Y., & Rao, Z. (2020). n-BiLSTM: BiLSTM with n-gram Features for Text Classification. *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 1056–1059. https://doi.org/10.1109/ITOEC49072.2020.9141692

Zhao, B. (2017). Web Scraping. Dalam *Encyclopedia of Big Data* (hlm. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_483-1

Zong, C., Xia, R., & Zhang, J. (2021). *Text Data Mining*. Springer Singapore. https://doi.org/10.1007/978-981-16-0100-2