

## DAFTAR PUSTAKA

- Abd, M., Mohammed, E., Al-Qaness, A. A., Ewees, A. A., dan Dahou, A. (2020). *Studies in Computational Intelligence 874 Recent Advances in NLP: The Case of Arabic Language*. <http://www.springer.com/series/7092>
- Agarwal, M. (2019). An Overview of Natural Language Processing. *International Journal for Research in Applied Science and Engineering Technology*, 7(5), 2811–2813. <https://doi.org/10.22214/ijraset.2019.5462>
- Al Amrani, Y., Lazaar, M., dan El Kadiri, K. E. (2018). Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis. *Procedia Computer Science*, 127, 511–520. <https://doi.org/10.1016/j.procs.2018.01.150>
- Aljameel, S. S., Alabbad, D. A., Alzahrani, N. A., Alqarni, S. M., Alamoudi, F. A., Babili, L. M., Aljaafary, S. K., dan Alshamrani, F. M. (2020). A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia. *International Journal of Environmental Research and Public Health*, 18(1), 218. <https://doi.org/10.3390/ijerph18010218>
- Badan Pusat Statistik. (2022). *Jumlah Pelanggan Perusahaan Air Bersih 2019-2021*. Badan Pusat Statistik. <https://www.bps.go.id/indicator/7/76/1/jumlah-pelanggan-perusahaan-air-bersih.html>
- Badan Pusat Statistik. (2023). *Jumlah Penduduk Pertengahan Tahun (ribu jiwa), 2020-2023*. Badan Pusat Statistik.
- Bagherzadeh, J., dan Asil, H. (2019). A review of various semi-supervised learning models with a deep learning and memory approach. *Iran Journal of Computer Science*, 2(2), 65–80. <https://doi.org/10.1007/s42044-018-00027-6>
- Bourequat, W., dan Mourad, H. (2021). Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine. *International*

*Journal of Advances in Data and Information Systems*, 2(1), 36–44.  
<https://doi.org/10.25008/ijadis.v2i1.1216>

Budi Raharjo. (2017). *Belajar Otodidak Flask (Framework Python untuk Pengembangan Aplikasi Web)*. Informatika Bandung.

Budi Raharjo. (2019). *Mudah Belajar Python Untuk Aplikasi Dekstop dan Web* (Vol. 2019). Informatika Bandung.

Chicco, D., dan Jurman, G. (2023). The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Mining*, 16(1), 4. <https://doi.org/10.1186/s13040-023-00322-4>

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bouusso, M., dan 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>

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., dan Romero, C. (2019). Text mining in education. Dalam *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 9, Nomor 6). Wiley-Blackwell. <https://doi.org/10.1002/widm.1332>

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., dan Romero, C. (2019). Text mining in education. *WIREs Data Mining and Knowledge Discovery*, 9(6). <https://doi.org/10.1002/widm.1332>

Fitriani, R. D., Yasin, H., dan Tarno, T. (2021). PENANGANAN KLASIFIKASI KELAS DATA TIDAK SEIMBANG DENGAN RANDOM OVERSAMPLING PADA NAIVE BAYES (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *Jurnal Gaussian*, 10(1), 11–20. <https://doi.org/10.14710/j.gauss.v10i1.30243>

Haiqal, H., Azmeri, A., dan Fatimah, E. (2020). Analisis Kinerja Pelayanan Air Bersih PDAM Tirta Daroy Kota Banda Aceh. *Jurnal Arsip Rekayasa Sipil dan Perencanaan*, 3(3), 203–210. <https://doi.org/10.24815/jarsp.v3i3.16643>

- Hairani, H., Saputro, K. E., dan Fadli, S. (2020). K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes. *Jurnal Teknologi dan Sistem Komputer*, 8(2), 89–93. <https://doi.org/10.14710/jtsiskom.8.2.2020.89-93>
- Hevner, A. R., March, S. T., Park, J., dan Ram, S. (2004). DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH 1. Dalam *Design Science in IS Research MIS Quarterly* (Vol. 28, Nomor 1).
- Hudaya, C. S., Fakhurroja, H., dan 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>
- Isnain, A. R., Sakti, A. I., Alita, D., dan Marga, N. S. (2021). SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM. *Jurnal Data Mining dan Sistem Informasi*, 2(1), 31. <https://doi.org/10.33365/jdmsi.v2i1.1021>
- Kacprzyk, J. (2019). *Studies in Big Data Series editor*. <http://www.springer.com/series/11970>
- Kar, C., Kumar, A., dan Banerjee, S. (2019). Tropical cyclone intensity detection by geometric features of cyclone images and multilayer perceptron. *SN Applied Sciences*, 1(9), 1099. <https://doi.org/10.1007/s42452-019-1134-8>
- Kasanah, A. N., Muladi, M., dan Pujianto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(2), 196–201. <https://doi.org/10.29207/resti.v3i2.945>
- Kementerian Dalam Negeri (Permendagri). (2007). *Peraturan Kementerian Dalam Negeri (Permendagri) Nomor 2 Tahun 2007*. <https://peraturan.bpk.go.id/Home/Details/126351/permendagri-no-2-tahun-2007>

- Kementrian Pekerjaan Umum dan Perumahan Rakyat. (2021). *Meski Semakin Langka, Air Tanah Masih Diminati Masyarakat*. Kementrian Pekerjaan Umum dan Perumahan Rakyat. <https://www.pu.go.id/index.php/berita/meski-semakin-langka-air-tanah-masih-diminati-masyarakat>
- Khan, Md. S. S., Rifa, S. R., Abir, A. E. H., dan Das, A. K. (2021). Sentiment Analysis on Bengali Facebook Comments To Predict Fan's Emotions Towards a Celebrity. *Journal of Engineering Advancements*, 118–124. <https://doi.org/10.38032/jea.2021.03.001>
- Kim, S.-W., dan Gil, J.-M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Human-centric Computing and Information Sciences*, 9(1), 30. <https://doi.org/10.1186/s13673-019-0192-7>
- Kubat, M. (2021). An Introduction to Machine Learning. Dalam *An Introduction to Machine Learning*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-81935-4>
- Lathkar, M. (2021). *Building web apps with Python and Flask learn to develop and deploy responsive RESTful web applications using Flask framework*. BPB Publications.
- Magfiroh, L., Sembiring, H., Sihombing, A., dan Kaputama Binjai, S. (2022). Clustering of Customer Complaints from PDAM Kota Binjai Using the K-Means Method. Dalam *International Journal Of Health, Engineering And Technology* (Vol. 1, Nomor 3). <https://ijhet.com/index.php/ijhess/>
- Mamgain, A. (2018). Guidance to Data Mining in Python. Dalam *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* © 2018 IJSRCSEIT / (Vol. 3, Nomor 10).
- Marcot, B. G., dan Hanea, A. M. (2021). What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Computational Statistics*, 36(3), 2009–2031. <https://doi.org/10.1007/s00180-020-00999-9>

- Moayedi, H., Osouli, A., Nguyen, H., dan Rashid, A. S. A. (2021). A novel Harris hawks' optimization and k-fold cross-validation predicting slope stability. *Engineering with Computers*, 37(1), 369–379. <https://doi.org/10.1007/s00366-019-00828-8>
- Mohan, L., Pant, J., Suyal, P., dan Kumar, A. (2020). Support Vector Machine Accuracy Improvement with Classification. *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, 477–481. <https://doi.org/10.1109/CICN49253.2020.9242572>
- Ncr, (, dan Clinton, J. (1999). *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler.
- Pahwa, B., Taruna, S., dan Kasliwal, N. (2018). Sentiment Analysis- Strategy for Text Pre-Processing. *International Journal of Computer Applications*, 180(34), 15–18. <https://doi.org/10.5120/ijca2018916865>
- 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., dan Duchesnay, E. (2011). *Scikit-learn: Machine Learning in Python*. [https://scikit-learn.org/stable/modules/feature\\_extraction.html#text-feature-extraction](https://scikit-learn.org/stable/modules/feature_extraction.html#text-feature-extraction)
- Persatuan Perusahaan Air Minum Seluruh Indonesia. (2023). *Persatuan Perusahaan Air Minum Seluruh Indonesia (Perpamsi)*. <https://perpamsi.or.id/anggota/biasa>
- Qaiser, S., dan Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181(1), 25–29. <https://doi.org/10.5120/ijca2018917395>
- Quinto, B. (2020). Next-generation machine learning with spark: Covers XGBoost, LightGBM, Spark NLP, distributed deep learning with keras, and more. Dalam *Next-Generation Machine Learning with Spark: Covers XGBoost, LightGBM, Spark NLP, Distributed Deep Learning with Keras, and More*. Apress Media LLC. <https://doi.org/10.1007/978-1-4842-5669-5>

- Rahmad, F., Suryanto, Y., dan Ramli, K. (2020). Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification. *IOP Conference Series: Materials Science and Engineering*, 879(1), 012076. <https://doi.org/10.1088/1757-899X/879/1/012076>
- Ridzuan, F., dan Wan Zainon, W. M. N. (2019). A Review on Data Cleansing Methods for Big Data. *Procedia Computer Science*, 161, 731–738. <https://doi.org/10.1016/j.procs.2019.11.177>
- Saabith, A. L. S., Fareez, M., dan Vinothraj, T. (2019). Related papers PYTHON CURRENT TREND APPLICATIONS-AN OVERVIEW POPULAR WEB DEVELOPMENT FRAMEWORKS IN PYTHON. *International Journal of Advance Engineering and Research Development*, 6(10).
- Santosa, B. (2007). Feature Selection with Support Vector Machines Applied on Tornado Detection. *IPTEK The Journal for Technology and Science*, 18(1). <https://doi.org/10.12962/j20882033.v18i1.178>
- Santosa, R. (2020). Quality of Public Service for Regional Water Companies: A Case Study in Local Water company Region II Makassar City. *International Journal of Multicultural and Multireligious Understanding*, 7(2), 498. <https://doi.org/10.18415/ijmmu.v7i2.1496>
- Santur, Y. (2019). Sentiment Analysis Based on Gated Recurrent Unit. *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1–5. <https://doi.org/10.1109/IDAP.2019.8875985>
- Thelwall, M., Buckley, K., dan Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1), 163–173. <https://doi.org/10.1002/asi.21662>
- Ustyannie, W., dan Suprpto, S. (2020). OVERSAMPLING METHOD TO HANDLING IMBALANCED DATASETS PROBLEM IN BINARY LOGISTIC REGRESSION ALGORITHM. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(1), 1. <https://doi.org/10.22146/ijccs.37415>

wearesocial. (2023). *The World's Most-Used Social Platforms*. wearesocial. <https://wearesocial.com/uk/blog/2023/01/the-changing-world-of-digital-in-2023/>

World Meteorological Organization. (2021). *Wake Up to The Looming Water Crisis, Report Warns*. World Meteorological Organization. <https://public.wmo.int/en/media/press-release/wake-looming-water-crisis-report-warns>

Wyatt, J. M., Booth, G. J., dan Goldman, A. H. (2021). Natural Language Processing and Its Use in Orthopaedic Research. *Current Reviews in Musculoskeletal Medicine*, 14(6), 392–396. <https://doi.org/10.1007/s12178-021-09734-3>

Xu, J., Zhang, Y., dan Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, 507, 772–794. <https://doi.org/10.1016/j.ins.2019.06.064>