DAFTAR PUSTAKA

Artzi, N. S., Shilo, S., Hadar, E., Rossman, H., Barbash-Hazan, S., Ben-Haroush, A., Balicer, R. D., Feldman, B., Wiznitzer, A., & Segal, E. (2020). Prediction of gestational diabetes based on nationwide electronic health records. *Nature Medicine*, *26*(1), 71–76. https://doi.org/10.1038/s41591-019-0724-8

Bilal, M., Ali, G., Iqbal, M., Anwar, M., Malik, M., & Kadir, R. (2022). Auto-Prep: Efficient and Automated Data Preprocessing Pipeline. *IEEE Access*, 10, 107764-107784. https://doi.org/10.1109/ACCESS.2022.3198662.

Bisong, E. (2019). Introduction to Scikit-learn. *Building machine learning and deep learning models on Google Cloud Platform: A comprehensive guide for beginners* (pp. 215–229). Apress. https://doi.org/10.1007/978-1-4842-4470-8_18

Bache-Mathiesen, L. K., Andersen, T. E., Dalen-Lorentsen, T., Clarsen, B., & Fagerland, M. W. (2021). Not straightforward: Modelling non-linearity in training load and injury research. *BMJ Open Sport & Exercise Medicine*, *7*(3), e001119. https://doi.org/10.1136/bmjsem-2021-001119

Chen, T., Xu, J., Ying, H., Chen, X., Feng, R., Fang, X., ... & Wu, J. (2019). Prediction of extubation failure for intensive care unit patients using light gradient boosting machine. *IEEE Access*, *7*, 150960-150968.

Cavus, M., & Biecek, P. (2022). Explainable expected goal models for performance analysis in football analytics. *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)* (Vol. 45, pp. 1–9). IEEE. https://doi.org/10.1109/dsaa54385.2022.10032440

Chatziparaskevas, P., Saprikis, V., & Antoniadis, I. (2024). The impact of information systems and data science on management in modern professional football: Moneyball theory and the development model of Brentford FC. *AIP Conference Proceedings*, *3220*, pp. 005-011. https://doi.org/10.1063/5.0237053

Das, L., Ahuja, L., Chauhan, V., & Singh, A. (2022). A review of data warehousing using feature engineering. *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, pp. 690–696. https://doi.org/10.1109/ICIPTM54933.2022.9754137

Eggels, H., Van Elk, R., & Pechenizkiy, M. (2016). Explaining soccer match outcomes with goal scoring opportunities predictive analytics. *3rd Workshop on Machine Learning and Data Mining for Sports Analytics (MLSA 2016)*. CEUR-WS.org

Fairchild, A., Pelechrinis, K., & Kokkodis, M. (2018). Spatial analysis of shots in MLS: A model for expected goals and fractal dimensionality. *Journal of Sports Analytics* (Vol. 4, Issue 3, pp. 165–174). IOS Press. https://doi.org/10.3233/jsa-170207

García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data *mining* . *Knowl. Based Syst.*, 98, 1-29. https://doi.org/10.1016/j.knosys.2015.12.006.

Hartanto, A. D., Nur Kholik, Y., & Pristyanto, Y. (2023). Stock Price Time Series Data Forecasting Using the Light Gradient Boosting Machine (LightGBM) Model. *JOIV: International Journal on Informatics Visualization, 7*(4), 2270–2279. https://doi.org/10.30630/joiv.7.4.1740

Hunt, J. (2019). Introduction to Matplotlib. *Advanced guide to Python 3 programming* (pp. 35–42). Springer International Publishing. https://doi.org/10.1007/978-3-030-25943-3_5

Katya, E. (2023). Exploring Feature Engineering Strategies for Improving Predictive Models in Data Science. *Research Journal of Computer Systems and Engineering*, *4*(2), 201–215. https://doi.org/10.52710/rjcse.88

Lucey, P., Bialkowski, A., Monfort, M., Carr, P., & Matthews, I. (2015). quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. *SPORTS ANALYTICS CONFERENCE, 9.* Disney Research. http://disneyresearch.s3.amazonaws.com

Mishra, V. N., Kumar, V., Prasad, R., & Punia, M. (2021). Geographically weighted method integrated with logistic regression for analyzing spatially varying

accuracy measures of remote sensing image classification. *Journal of the Indian Society of Remote Sensing*, *49*(5), 1189–1199. https://doi.org/10.1007/s12524-020-01286-2

Molin, S., & Jee, K. (2021). *Hands-on data analysis with Pandas: A Python data science handbook for data collection, wrangling, analysis, and visualization*. Packt Publishing.

Nelli, F. (2015). pandas: Reading and writing data. *Python data analytics: Data analysis and science using pandas, matplotlib, and the Python programming language* (pp. 103–130). Apress. https://doi.org/10.1007/978-1-4842-0958-5_5

Nemeth, M., Borkin, D., & Michalconok, G. (2019). The comparison of Machine-Learning Methods XGBOOST and LightGBM to predict energy development. In *Advances in intelligent systems and computing* (pp. 208–215). https://doi.org/10.1007/978-3-030-31362-3_21

Pérez, F., Granger, B. E., & Hunter, J. D. (2011). Python: An ecosystem for scientific computing. *Computing in Science & Engineering, 13*(2), 13–21. https://doi.org/10.1109/MCSE.2010.119

Pratama, J. A., Suprijadi, Y., & Zulhanif, Z. (2017). The Analisis Sentimen Sosial Media Twitter Dengan Algoritma Machine Learning Menggunakan Software R. *Jurnal Fourier* (Vol. 6, Issue 2, p. 85). Al-Jamiah Research Centre. https://doi.org/10.14421/fourier.2017.62.85-89

Ramadanti, E. ., Aprilya Dinathi, D. ., christianskaditya, & Chandranegara, D. R. . (2024). Diabetes Disease Detection Classification Using Light Gradient Boosting (LightGBM) With Hyperparameter Tuning. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *8*(2), 956-963. https://doi.org/10.33395/sinkron.v8i2.13530

Idris, H. I., Mohammed, A., Salisu, U. F., Balansana, K. I., Abdulazeez, D., & Danrimi, N. H. (2024). Evaluating the performances of robust logistic regression models in the presence of outliers. *African Journal of Mathematics and Statistics Studies*, *7*(4), 320–327. https://doi.org/10.52589/AJMSS-YKDFCYQS

Sammut, C., & Webb, G. I. (2017). Data preprocessing. *Encyclopedia of machine learning and data mining* (p. 327). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_100100

Srinath, K.R. (2017). Python – The Fastest Growing Programming Language. *International Research Journal of Engineering and Technology (IRJET), 4*(12), 354-357.

Sutton, C., & McCallum, A. (2012). *An introduction to conditional random fields*. *Foundations and Trends® in Machine Learning*, *4*(4), 267–373. https://doi.org/10.1561/2200000013

Statista. (2024). *Soccer - Indonesia | Statista market forecast*. Retreived October 27, 2024, from https://www.statista.com/outlook/amo/sports/soccer/indonesia

StatsBomb. (2022). *StatsBomb open data*. GitHub. Retreived October 28, 2024, from https://github.com/statsbomb/open-data

StatsBomb. (2024). *Who are we?.* Retreived October 28, 2024, from https://statsbomb.com/who-we-are

Tureen, T., & Olthof, S. (2022). Estimated Player Impact (EPI): Quantifying the effects of individual players on football (soccer) actions using hierarchical statistical models. *StatsBomb Conference 2022*. 9.

Verdonck, T., Baesens, B., Óskarsdóttir, M., & Broucke, S. (2021). Special issue on feature engineering editorial. *Mach Learn*, *113*(7), 3917-3928. https://doi.org/10.1007/s10994-021-06042-2.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software, 6*(60), 1–4. https://doi.org/10.21105/joss.03021

Whitmore, J. (2023). *What is expected goals (xG)?* Opta Analyst. Retrieved October 27, 2024, from https://theanalyst.com/eu/2023/08/what-is-expected-goals-xg/

McCulloch, C. E., & Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. Statistical Science, 26(3), 388–402. https://doi.org/10.1214/11-STS361

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. Trends in Ecology & Evolution, 24(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Sheridan, R. P., Liaw, A., & Tudor, M. (2021). Light Gradient Boosting Machine as a regression method for quantitative structure-activity relationships. *arXiv preprint*. https://arxiv.org/abs/2105.08626