

DAFTAR PUSTAKA

- Adeniyi, A. E., Madamidola, O. A., Awotunde, J. B., Misra, S., & Agrawal, A. (2024). Comparative Analysis of CNN and SVM Machine Learning Techniques for Plant Disease Detection. Dalam J. Agrawal, R. K. Shukla, S. Sharma, & C.-S. Shieh (Ed.), *Data Engineering and Applications* (Vol. 1146, hlm. 389–402). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-0037-0_30
- Aldabbagh, A. D. A., Hairu, C., & Hanafi, M. (2020). Classification of Chili Plant Growth using Deep Learning. *2020 IEEE 10th International Conference on System Engineering and Technology (ICSET)*, 213–217. <https://doi.org/10.1109/ICSET51301.2020.9265351>
- Alexandropoulos, S.-A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. *The Knowledge Engineering Review*, 34, e1. <https://doi.org/10.1017/S026988891800036X>
- Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, 21(11), 3758. <https://doi.org/10.3390/s21113758>
- Bhagat, M., & Kumar, D. (2023). Performance evaluation of PCA based reduced features of leaf images extracted by DWT using random Forest and XGBoost classifier. *Multimedia Tools and Applications*, 82(17), 26225–26254. <https://doi.org/10.1007/s11042-023-14370-9>
- Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70–79. <https://doi.org/10.1016/j.neucom.2017.11.077>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chicco, D., & 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>

- Corrales, D. C., Department of Computer Science and Engineering, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911, Leganés, Spain, & Corrales, J. C. (2015). A Conceptual Framework for Data Quality in Knowledge Discovery Tasks (FDQ-KDT): A Proposal. *Journal of Computers*, 10(6), 396–405. <https://doi.org/10.17706/jcp.10.6.396-405>
- Côté, P.-O., Nikanjam, A., Ahmed, N., Humeniuk, D., & Khomh, F. (2024). Data cleaning and machine learning: A systematic literature review. *Automated Software Engineering*, 31(2), 54. <https://doi.org/10.1007/s10515-024-00453-w>
- Delai, C., Muhae-Ud-Din, G., Abid, R., Tian, T., Liu, R., Xiong, Y., Ma, S., & Ghorbani, A. (2024). A comprehensive review of integrated management strategies for damping-off disease in chili. *Frontiers in Microbiology*, 15, 1479957. <https://doi.org/10.3389/fmicb.2024.1479957>
- Fakhrurroja, H., Rofi Hidayatullah, A., Pramesti, D., & Ismail, N. (2024). Classification of Diseases in Chili Plants Using the SVM Method: Development and Implementation. *2024 12th International Conference on Information and Communication Technology (ICoICT)*, 493–499. <https://doi.org/10.1109/ICoICT61617.2024.10698555>
- Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, 10, 19083–19095. <https://doi.org/10.1109/ACCESS.2022.3151048>
- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematics*, 10(11), 1942. <https://doi.org/10.3390/math10111942>
- Jeong, S., Kim, Y., Hur, S. H., Bang, H., Kim, H., & Chung, H. (2023). Explainable extreme gradient boosting as a machine learning tool for discrimination of the geographical origin of chili peppers using laser ablation–inductively coupled plasma mass spectrometry, X-ray fluorescence, and near-infrared spectroscopy. *Journal of Agriculture and Food Research*, 18. <https://doi.org/10.1016/j.jafr.2024.101446>
- Karim, K. M. R., Rafii, M. Y., Misran, A. B., Ismail, M. F. B., Harun, A. R., Khan,

- M. M. H., & Chowdhury, Mst. F. N. (2021). Current and Prospective Strategies in the Varietal Improvement of Chilli (*Capsicum annuum* L.) Specially Heterosis Breeding. *Agronomy*, *11*(11), 2217. <https://doi.org/10.3390/agronomy11112217>
- Karim, K. M. R., Rafii, M. Y., Misran, A., Ismail, M. F., Harun, A. R., Ridzuan, R., Chowdhury, Mst. F. N., Hosen, M., Yusuff, O., & Haque, M. A. (2022). Genetic Diversity Analysis among *Capsicum annuum* Mutants Based on Morpho-Physiological and Yield Traits. *Agronomy*, *12*(10), 2436. <https://doi.org/10.3390/agronomy12102436>
- Lathkar, M. (2021). *Building Web Apps with Python and Flask* (First Edition). BPB Publications.
- Lemaitre, G., & Nogueira, F. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, *18*(1), 1–5.
- Li, Y., Stasinakis, C., & Yeo, W. M. (2022). A Hybrid XGBoost-MLP Model for Credit Risk Assessment on Digital Supply Chain Finance. *Forecasting*, *4*(1), 184–207. <https://doi.org/10.3390/forecast4010011>
- Mariyono, J., & Sumarno, S. (2015). Chilli production and adoption of chilli-based agribusiness in Indonesia. *Journal of Agribusiness in Developing and Emerging Economies*, *5*(1), 57–75. <https://doi.org/10.1108/JADEE-01-2014-0002>
- Marlina, I Komang Damar Jaya, & Jayaputra. (2024). Pertumbuhan Dan Hasil Dua Varietas Cabai Merah (*Capsicum annum* L.) Di Lahan Kering Yang Ditumpangsarikan dengan Tanaman Kacang Tanah Pada Waktu Tanam Berbeda. *Jurnal Ilmiah Mahasiswa Agrokomplek*, *3*(2), 45–52. <https://doi.org/10.29303/jima.v3i2.4671>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *33*(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>

- M'hamdi, O., Takács, S., Palotás, G., Ilahy, R., Helyes, L., & Pék, Z. (2024). A Comparative Analysis of XGBoost and Neural Network Models for Predicting Some Tomato Fruit Quality Traits from Environmental and Meteorological Data. *Plants*, 13(5), 746. <https://doi.org/10.3390/plants13050746>
- Moon, T., Park, J., & Son, J. E. (2021). Prediction of the fruit development stage of sweet pepper (*Capsicum annum* var. *Annuum*) by an ensemble model of convolutional and multilayer perceptron. *Biosystems Engineering*, 210, 171–180. <https://doi.org/10.1016/j.biosystemseng.2021.08.017>
- Muflikh, Y. N., Smith, C., Brown, C., & Aziz, A. A. (2021). Analysing price volatility in agricultural value chains using systems thinking: A case study of the Indonesian chilli value chain. *Agricultural Systems*, 192, 103179. <https://doi.org/10.1016/j.agsy.2021.103179>
- Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, 2021, 1–15. <https://doi.org/10.1155/2021/4832864>
- Omari Firas. (2023). A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis based knowledge discovery: A data mining approach. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 009–014. <https://doi.org/10.30574/wjaets.2023.8.1.0147>
- Padilha, H. K. M., Sigales, C. V., Villela, J. C. B., Valgas, R. A., & Lía, R. (2016). Agronomic evaluation and morphological characterization of chili peppers (*Capsicum annum*, Solanaceae) from Brazil. *Australian Journal of Basic and Applied Sciences*.
- Palacios, H. J. G., Toledo, R. A. J., Pantoja, G. A. H., & Navarro, Á. A. M. (2017). A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change. *Advances in Science, Technology and Engineering Systems Journal*, 2(3), 598–604. <https://doi.org/10.25046/aj020376>

- Pinheiro, J. M. H., Oliveira, S. V. B. de, Silva, T. H. S., Saraiva, P. A. R., Souza, E. F. de, Ambrosio, L. A., & Becker, M. (2025). *The Impact of Feature Scaling In Machine Learning: Effects on Regression and Classification Tasks* (No. arXiv:2506.08274). arXiv. <https://doi.org/10.48550/arXiv.2506.08274>
- Putra Laksamana, A. D., Fakhrurroja, H., & Pramesti, D. (2024). Developing a Labeled Dataset for Chili Plant Health Monitoring: A Multispectral Image Segmentation Approach with YOLOv8. *2024 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 440–445. <https://doi.org/10.1109/IC3INA64086.2024.10732221>
- Rahevar, P. M., Patel, J. N., Kumar, S., & Acharya, R. R. (2019). Morphological, biochemical and molecular characterization for genetic variability analysis of *Capsicum annuum*. *Vegetos*, 32(2), 131–141. <https://doi.org/10.1007/s42535-019-00016-5>
- Rahman, A. (2019). Statistics-Based Data Preprocessing Methods and Machine Learning Algorithms for Big Data Analysis. *International Journal of Artificial Intelligence*, 17(2), 44–65.
- Rakhman, A., Sutejo, H., & Jannah, N. (2023). Respon Pertumbuhan dan Hasil Tanaman Cabai Besar (*Capsicum annuum* L.) Varietas Baja F1 Terhadap Pemberian Pupuk Petroganik dan SP-36. *JAKT : Jurnal Agroteknologi dan Kehutanan Tropika*, 1(1), 1. <https://doi.org/10.31293/jakt.v1i1.6619>
- Relan, K. (2019). *Building REST APIs with Flask: Create Python Web Services with MySQL*. Apress. <https://doi.org/10.1007/978-1-4842-5022-8>
- Rotondo, A., & Quilligan, F. (2020). Evolution Paths for Knowledge Discovery and Data Mining Process Models. *SN Computer Science*, 1(2), 109. <https://doi.org/10.1007/s42979-020-0117-6>
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Showrav, T. T., Bain, S., Hossain, M., Ahmed, K. I., Fattah, S. A., & Ahmed, S. (2022). A Two-stage Approach for Plant Disease Classification Based on Deep Neural Networks and Transfer Learning. *2022 12th International*

- Conference on Electrical and Computer Engineering (ICECE)*, 469–472.
<https://doi.org/10.1109/ICECE57408.2022.10088587>
- Simonetto, T., Ghamizi, S., & Cordy, M. (2024). *TabularBench: Benchmarking Adversarial Robustness for Tabular Deep Learning in Real-world Use-cases* (No. arXiv:2408.07579). arXiv.
<https://doi.org/10.48550/arXiv.2408.07579>
- Sivaranjani, A., Senthilrani, S., Ashok Kumar, B., & Senthil Murugan, A. (2022). An Overview of Various Computer Vision-based Grading System for Various Agricultural Products. *The Journal of Horticultural Science and Biotechnology*, 97(2), 137–159.
<https://doi.org/10.1080/14620316.2021.1970631>
- Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., & Dosovitskiy, A. (2021). *MLP-Mixer: An all-MLP Architecture for Vision*.
- Valero-Carreras, D., Alcaraz, J., & Landete, M. (2023). Comparing two SVM models through different metrics based on the confusion matrix. *Computers & Operations Research*, 152, 106131.
<https://doi.org/10.1016/j.cor.2022.106131>
- Wongvorachan, T., He, S., & Bulut, O. (2023). A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information*, 14(1), 54.
<https://doi.org/10.3390/info14010054>
- Xiong, Z., Cui, Y., Liu, Z., Zhao, Y., Hu, M., & Hu, J. (2020). Evaluating explorative prediction power of machine learning algorithms for materials discovery using k -fold forward cross-validation. *Computational Materials Science*, 171, 109203.
<https://doi.org/10.1016/j.commatsci.2019.109203>
- Zhang, H., Li, W., Wang, G., Song, F., Wen, Z., Zhang, H., Tong, L., & Kang, S. (2025). Predicting stomatal conductance of chili peppers using TPE-optimized LightGBM and SHAP feature analysis based on UAVs' hyperspectral, thermal infrared imagery, and meteorological data.

Computers and Electronics in Agriculture, 231, 110036.
<https://doi.org/10.1016/j.compag.2025.110036>

Zhang, P., Jia, Y., & Shang, Y. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, 18(6), 155013292211069. <https://doi.org/10.1177/15501329221106935>

Zhao, W., Lai, X., Liu, D., Zhang, Z., Ma, P., Wang, Q., Zhang, Z., & Pan, Y. (2020). Applications of Support Vector Machine in Genomic Prediction in Pig and Maize Populations. *Frontiers in Genetics*, 11, 598318. <https://doi.org/10.3389/fgene.2020.598318>