

DAFTAR PUSTAKA

- Aized, S., & Arshad, A. (2017). Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*, 13(August), 459–465. <https://doi.org/10.6000/1927-5129.2017.13.76>
- Ammaridho, M. (2022). Implementasi machine learning dengan metode algoritma decision tree c4. 5 untuk pemilihan bandwidth internet wifi rumahan. In *Repository.Uinjkt.Ac.Id*. https://repository.uinjkt.ac.id/dspace/handle/123456789/65304%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/65304/1/MUHAMMAD_AMMARIDHO_R._FST.pdf
- Azizah, N. (2021). Komparasi Metode Klasifikasi Decision Tree Algoritma C4.5 Dan Random Forest Untuk Prediksi Penyakit Stroke. *Skripsi*.
- Bhavsar, H., & Ganatra, A. (2012). A Comparative Study of Training Algorithms for Supervised Machine Learning. *International Journal of Soft Computing and Engineering*, 2(4), 74–81.
- Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*, 8(1), 884–898.
- Cao, H., Li, X. L., Woon, D. Y. K., & Ng, S. K. (2013). Integrated oversampling for imbalanced time series classification. *IEEE Transactions on Knowledge and Data Engineering*, 25(12), 2809–2822. <https://doi.org/10.1109/TKDE.2013.37>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). snopes.com: Two-Striped Telamonia Spider. *Journal of Artificial Intelligence Research*, 16(Sept. 28), 321–357. <https://arxiv.org/pdf/1106.1813.pdf%0Ahttp://www.snopes.com/horrors/insects/telamonia.asp>
- Debie, E., & Shafi, K. (2019). Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. *Pattern Analysis and Applications*, 22(2), 519–536. <https://doi.org/10.1007/s10044-017-0649-0>
- Demir, Y., & Hassan, M. (2021). Combination of PCA with SMOTE Oversampling for Classification of High-Dimensional Imbalanced Data. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, September*. <https://doi.org/10.17798/bitlisfen.939733>
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques*.
- Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis.: Pearson Prentice Hall. In *Pearson Prentice Hall*.
- Jolliffe, I. T. (2002). Principal components analysis. *International Encyclopedia of Education*, 374–377. <https://doi.org/10.1016/B978-0-08-044894-7.01358-0>
- Khadijah, K., & Sasongko, P. S. (2020). The Comparison of Imbalanced Data Handling Method in Software Defect Prediction. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(3), 203–210. <https://doi.org/10.22219/kinetik.v5i3.1049>
- Laradji, I. H., Alshayeb, M., & Ghouti, L. (2015). Software defect prediction using ensemble learning on selected features. *Information and Software Technology*, 58(July), 388–402. <https://doi.org/10.1016/j.infsof.2014.07.005>
- Latifah, R., Wulandari, E. S., & Kreshna, P. E. (2019). Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn. *Jurnal Universitas Muhammadiyah Jakarta*, 1–6. <https://jurnal.umj.ac.id/index.php/semnastek/article/download/5239/3517>
- Liu, T., Fan, W., & Wu, C. (2019). A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial Intelligence in Medicine*, 101,

101723. <https://doi.org/10.1016/j.artmed.2019.101723>
- Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th International Conference on Information and Communication Systems, ICICS 2020, April*, 243–248. <https://doi.org/10.1109/ICICS49469.2020.239556>
- Mutmainah, S. (2021). Penanganan Imbalance Data Pada Klasifikasi. *SNATi*, 1, 10–16.
- Naseriparsa, M., & Mansour, M. (2013). *Combination of PCA with SMOTE Resampling to Boost the Prediction Rate in Lung Cancer Dataset*. 77(3), 33–38.
- Nugroho, F. (2019). *Sistem Rekomendasi Kata Kunci Untuk Website Menggunakan Hybrid Semantic Relatedness Dan Associative Neural Network*. 6–27.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21–44. <https://doi.org/10.1109/MCAS.2006.1688199>
- Pristyanto, Y. (2019). Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset. *Jurnal Teknoinfo*, 13(1), 11. <https://doi.org/10.33365/jti.v13i1.184>
- Rofiq, H., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 3(1), 8–15. <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19417.pdf>
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1–2), 1–39. <https://doi.org/10.1007/s10462-009-9124-7>
- Saifudin, A., & Wahono, R. S. (2015). Penerapan Teknik Ensemble untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *Journal of Software Engineering*, 1(1).
- Salsabila, M. (2022). *Pendekatan Visual Analytics Dalam Pemodelan Prediksi Cacat Perangkat Lunak Menggunakan Kombinasi Pca Dan Smote*.
- Singh, P., & Chug, A. (2019). Software defect prediction analysis by using machine learning algorithms. *International Journal of Recent Technology and Engineering*, 8(2 Special Issue 11), 3544–3546. <https://doi.org/10.35940/ijrte.B1438.0982S1119>
- Stein, G., Wu, A. S., & Hua, K. A. (2005). *10.1.1.80.1400.Pdf*.
- Sulistiyono, M., Pristyanto, Y., Adi, S., & Gumelar, G. (2021). Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi. *Sistemasi*, 10(2), 445. <https://doi.org/10.32520/stmsi.v10i2.1303>
- Topîrceanu, A., & Grossecck, G. (2017). Decision tree learning used for the classification of student archetypes in online courses. *Procedia Computer Science*, 112, 51–60. <https://doi.org/10.1016/j.procs.2017.08.021>
- Velicer, W. F., & Molenaar, P. C. (2012). Time Series Analysis for Psychological Research. In *Handbook of Psychology, Second Edition* (Issue April 2003). <https://doi.org/10.1002/9781118133880.hop202022>
- Wahyu, A. (2016). *Perbandingan Kinerja AUC dan G-Means Pada Machine Learning Berbasis Seleksi Fitur Algoritma Genetika Untuk Prediksi Cacat Perangkat Lunak*. 1–23.
- Wang, H., Xu, P., & Zhao, J. (2022). Improved KNN algorithms of spherical regions based on clustering and region division. *Alexandria Engineering Journal*, 61(5), 3571–3585. <https://doi.org/10.1016/j.aej.2021.09.004>

- Wardani, N. W. (2020). Penerapan Data Mining dalam Analytic CRM. In *Yayasan Kita Menulis* (Issue May). <https://www.researchgate.net/publication/351776255>
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1(2), 56–70. <https://doi.org/10.38094/jastt1224>
- Aized, S., & Arshad, A. (2017). Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*, 13(August), 459–465. <https://doi.org/10.6000/1927-5129.2017.13.76>
- Ammaridho, M. (2022). Implementasi machine learning dengan metode algoritma decision tree c4. 5 untuk pemilihan bandwidth internet wifi rumahan. In *Repository.Uinjkt.Ac.Id*. [https://repository.uinjkt.ac.id/dspace/handle/123456789/65304%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/65304/1/MUHAMMAD AMMARIDHO R. S.-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/65304%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/65304/1/MUHAMMAD_AMMARIDHO_R._S.-FST.pdf)
- Azizah, N. (2021). Komparasi Metode Klasifikasi Decision Tree Algoritma C4.5 Dan Random Forest Untuk Prediksi Penyakit Stroke. *Skripsi*.
- Bhavsar, H., & Ganatra, A. (2012). A Comparative Study of Training Algorithms for Supervised Machine Learning. *International Journal of Soft Computing and Engineering*, 2(4), 74–81.
- Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*, 8(1), 884–898.
- Cao, H., Li, X. L., Woon, D. Y. K., & Ng, S. K. (2013). Integrated oversampling for imbalanced time series classification. *IEEE Transactions on Knowledge and Data Engineering*, 25(12), 2809–2822. <https://doi.org/10.1109/TKDE.2013.37>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). snopes.com: Two-Striped Telamonia Spider. *Journal of Artificial Intelligence Research*, 16(Sept. 28), 321–357. <https://arxiv.org/pdf/1106.1813.pdf%0Ahttp://www.snopes.com/horrors/insects/telamonia.asp>
- Debie, E., & Shafi, K. (2019). Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. *Pattern Analysis and Applications*, 22(2), 519–536. <https://doi.org/10.1007/s10044-017-0649-0>
- Demir, Y., & Hassan, M. (2021). Combination of PCA with SMOTE Oversampling for Classification of High-Dimensional Imbalanced Data. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, September. <https://doi.org/10.17798/bitlisfen.939733>
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques*.
- Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis.: Pearson Prentice Hall. In *Pearson Prentice Hall*.
- Jolliffe, I. T. (2002). Principal components analysis. *International Encyclopedia of Education*, 374–377. <https://doi.org/10.1016/B978-0-08-044894-7.01358-0>
- Khadijah, K., & Sasongko, P. S. (2020). The Comparison of Imbalanced Data Handling Method in Software Defect Prediction. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(3), 203–210. <https://doi.org/10.22219/kinetik.v5i3.1049>
- Laradji, I. H., Alshayeb, M., & Ghouti, L. (2015). Software defect prediction using ensemble learning on selected features. *Information and Software Technology*, 58(July), 388–402. <https://doi.org/10.1016/j.infsof.2014.07.005>

- Latifah, R., Wulandari, E. S., & Kreshna, P. E. (2019). Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn. *Jurnal Universitas Muhammadiyah Jakarta*, 1–6. <https://jurnal.umj.ac.id/index.php/semnastek/article/download/5239/3517>
- Liu, T., Fan, W., & Wu, C. (2019). A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial Intelligence in Medicine*, 101, 101723. <https://doi.org/10.1016/j.artmed.2019.101723>
- Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th International Conference on Information and Communication Systems, ICICS 2020, April*, 243–248. <https://doi.org/10.1109/ICICS49469.2020.239556>
- Mutmainah, S. (2021). Penanganan Imbalance Data Pada Klasifikasi. *SNATi*, 1, 10–16.
- Naseriparsa, M., & Mansour, M. (2013). *Combination of PCA with SMOTE Resampling to Boost the Prediction Rate in Lung Cancer Dataset*. 77(3), 33–38.
- Nugroho, F. (2019). *Sistem Rekomendasi Kata Kunci Untuk Website Menggunakan Hybrid Semantic Relatedness Dan Associative Neural Network*. 6–27.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21–44. <https://doi.org/10.1109/MCAS.2006.1688199>
- Pristyanto, Y. (2019). Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset. *Jurnal Teknoinfo*, 13(1), 11. <https://doi.org/10.33365/jti.v13i1.184>
- Rofiq, H., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 3(1), 8–15. <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19417.pdf>
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1–2), 1–39. <https://doi.org/10.1007/s10462-009-9124-7>
- Saifudin, A., & Wahono, R. S. (2015). Penerapan Teknik Ensemble untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *Journal of Software Engineering*, 1(1).
- Salsabila, M. (2022). *Pendekatan Visual Analytics Dalam Pemodelan Prediksi Cacat Perangkat Lunak Menggunakan Kombinasi Pca Dan Smote*.
- Singh, P., & Chug, A. (2019). Software defect prediction analysis by using machine learning algorithms. *International Journal of Recent Technology and Engineering*, 8(2 Special Issue 11), 3544–3546. <https://doi.org/10.35940/ijrte.B1438.0982S1119>
- Stein, G., Wu, A. S., & Hua, K. A. (2005). *10.1.1.80.1400.Pdf*.
- Sulistiyono, M., Pristyanto, Y., Adi, S., & Gumelar, G. (2021). Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi. *Sistemasi*, 10(2), 445. <https://doi.org/10.32520/stmsi.v10i2.1303>
- Topîrceanu, A., & Grossecck, G. (2017). Decision tree learning used for the classification of student archetypes in online courses. *Procedia Computer Science*, 112, 51–60. <https://doi.org/10.1016/j.procs.2017.08.021>
- Velicer, W. F., & Molenaar, P. C. (2012). Time Series Analysis for Psychological Research. In *Handbook of Psychology, Second Edition* (Issue April 2003). <https://doi.org/10.1002/9781118133880.hop202022>

- Wahyu, A. (2016). *Perbandingan Kinerja AUC dan G-Means Pada Machine Learning Berbasis Seleksi Fitur Algoritma Genetika Untuk Prediksi Cacat Perangkat Lunak*. 1–23.
- Wang, H., Xu, P., & Zhao, J. (2022). Improved KNN algorithms of spherical regions based on clustering and region division. *Alexandria Engineering Journal*, 61(5), 3571–3585. <https://doi.org/10.1016/j.aej.2021.09.004>
- Wardani, N. W. (2020). Penerapan Data Mining dalam Analytic CRM. In *Yayasan Kita Menulis* (Issue May). <https://www.researchgate.net/publication/351776255>
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1(2), 56–70. <https://doi.org/10.38094/jastt1224>