

DAFTAR PUSTAKA

- [1] PCMI, “A Comprehensive Look at Indonesia’s E-Commerce Market,” Payments and Commerce Market Intelligence. Accessed: Jan. 27, 2026. [Online]. Available: <https://paymentscmi.com/insights/indonesia-ecommerce-market-data/>
- [2] Matteo Tittarelli, “Customer Acquisition Cost Benchmarks - 44 Statistics Every Marketing Leader Should Know in 2025,” Genesys. Accessed: Jan. 27, 2026. [Online]. Available: <https://genesysgrowth.com/blog/customer-acquisition-cost-benchmarks-for-marketing-leaders>
- [3] K. Hasumoto and M. Goto, “Predicting customer churn for platform businesses: using latent variables of variational autoencoder as consumers’ purchasing behavior,” *Neural Comput. Appl.*, vol. 34, no. 21, pp. 18525–18541, Nov. 2022, doi: 10.1007/s00521-022-07418-8.
- [4] P. S. Kumar, “Predicting Customer Churn in E-Commerce Using Machine Learning: A Comprehensive Approach,” *IJIIS: International Journal of Informatics and Information Systems*, vol. 9, no. 2, pp. 357–370, Mar. 2026, doi: 10.47738/ijiis.v9i2.300.
- [5] M. Fridrich and P. Dostál, “User Churn Model in E-Commerce Retail,” *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration*, vol. 30, no. 1, 2022, doi: 10.46585/SP30011478.
- [6] R. Suguna, J. Suriya Prakash, H. Aditya Pai, T. R. Mahesh, V. Vinoth Kumar, and T. E. Yimer, “Mitigating class imbalance in churn prediction with ensemble methods and SMOTE,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-01031-0.
- [7] P. Boozary, S. Sheykhan, H. GhorbanTanhaei, and C. Magazzino, “Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction,” *International Journal of Information Management Data Insights*, vol. 5, no. 1, Jun. 2025, doi: 10.1016/j.jjimei.2025.100331.
- [8] N. Ahmad, M. J. Awan, H. Nobanee, A. M. Zain, A. Naseem, and A. Mahmoud, “Customer Personality Analysis for Churn Prediction Using Hybrid Ensemble Models and Class Balancing Techniques,” *IEEE Access*, vol. 12, pp. 1865–1879, 2024, doi: 10.1109/ACCESS.2023.3334641.
- [9] A. Manzoor, M. Atif Qureshi, E. Kidney, and L. Longo, “A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners,” *IEEE Access*, vol. 12, pp. 70434–70463, 2024, doi: 10.1109/ACCESS.2024.3402092.
- [10] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, “Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for

evaluating business predictive models,” *J. Big Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1186/s40537-025-01313-4.

- [11] L. Matlala, T. Bokaba, P. Ndayizigamiye, S. Mhlongo, and E. Dogo, “A comparative analysis of ensemble learning models for predicting lapses in investment policies,” *Journal of Management Analytics*, 2025, doi: 10.1080/23270012.2025.2574030.
- [12] R. E. Ako *et al.*, “Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost,” *Journal of Computing Theories and Applications*, vol. 2, no. 1, pp. 86–101, Jun. 2024, doi: 10.62411/jcta.10562.
- [13] R. Bhuria *et al.*, “Ensemble-based customer churn prediction in banking: a voting classifier approach for improved client retention using demographic and behavioral data,” *Discover Sustainability*, vol. 6, no. 1, Dec. 2025, doi: 10.1007/s43621-025-00807-8.
- [14] M. Munzir, R. Mubarak, Y. H. Chrisnanto, and P. N. Sabrina, “Enrichment: Journal of Multidisciplinary Research and Development Implementation of Random Forest Using Smote and Smoteenn in Customer Churn Classification in E-Commerce,” Nov. 2023. doi: 10.55324/enrichment.v1i8.69.
- [15] K. Matuszelański and K. Kopczewska, “Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 1, pp. 165–198, Mar. 2022, doi: 10.3390/jtaer17010009.
- [16] S. J. C. Gangadhar, R. K. Arora, P. N. Renjith, J. Bamini, and Y. devidas Chincholkar, “E-commerce customer churn prevention using machine learning-based business intelligence strategy,” *Measurement: Sensors*, vol. 27, Jun. 2023, doi: 10.1016/j.measen.2023.100728.
- [17] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, “Customer churn prediction using composite deep learning technique,” *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-44396-w.
- [18] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, pp. 685–695, Apr. 2021, doi: 10.1007/s12525-021-00475-2/Published.
- [19] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” May 01, 2021, *Springer*. doi: 10.1007/s42979-021-00592-x.
- [20] A. A. Khan, O. Chaudhari, and R. Chandra, “A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation,” Jun. 15, 2024, *Elsevier Ltd*. doi: 10.1016/j.eswa.2023.122778.
- [21] S. Wang, Y. Dai, J. Shen, and J. Xuan, “Research on expansion and classification of imbalanced data based on SMOTE algorithm,” *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-03430-5.

- [22] P. Thölke *et al.*, “Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data,” *Neuroimage*, vol. 277, Aug. 2023, doi: 10.1016/j.neuroimage.2023.120253.
- [23] H. Jafarzadeh, M. Mahdianpari, E. Gill, F. Mohammadimanesh, and S. Homayouni, “Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and polSAR data: A comparative evaluation,” *Remote Sens. (Basel)*, vol. 13, no. 21, Nov. 2021, doi: 10.3390/rs13214405.
- [24] D. Tiwari, B. Nagpal, B. S. Bhati, A. Mishra, and M. Kumar, “A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques,” *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13407–13461, Nov. 2023, doi: 10.1007/s10462-023-10472-w.
- [25] I. D. Mienye and Y. Sun, “A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects,” 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3207287.
- [26] A. Satoła and K. Satoła, “Performance comparison of machine learning models used for predicting subclinical mastitis in dairy cows: Bagging, boosting, stacking, and super-learner ensembles versus single machine learning models,” *J. Dairy Sci.*, vol. 107, no. 6, pp. 3959–3972, Jun. 2024, doi: 10.3168/jds.2023-24243.
- [27] L. T. Tam, L. G. Vi, and N. M. Tuan, “Comparison of Methods for Handling Imbalanced Data in Customer Churn Prediction with Feature Selection Using SHAP and mRMR Frameworks,” *Cybernetics and Information Technologies*, vol. 25, no. 3, pp. 68–87, Sep. 2025, doi: 10.2478/cait-2025-0023.
- [28] M. Imani, A. Beikmohammadi, and H. R. Arabnia, “Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels,” *Technologies (Basel)*, vol. 13, no. 3, Mar. 2025, doi: 10.3390/technologies13030088.
- [29] P. S. Kumar, “Predicting Customer Churn in E-Commerce Using Machine Learning: A Comprehensive Approach,” *IJIIS: International Journal of Informatics and Information Systems*, vol. 9, no. 2, pp. 357–370, Mar. 2026, doi: 10.47738/ijiis.v9i2.300.
- [30] J. P. Bharti, P. Mishra, U. moorthy, V. E. Sathishkumar, Y. Cho, and P. Samui, “Slope Stability Analysis Using Rf, Gbm, Cart, Bt and Xgboost,” *Geotechnical and Geological Engineering*, vol. 39, no. 5, pp. 3741–3752, Jun. 2021, doi: 10.1007/s10706-021-01721-2.
- [31] Rahmanul Hoque, Suman Das, Mahmudul Hoque, and Mahmudul Hoque, “Breast Cancer Classification using XGBoost,” *World Journal of Advanced Research and Reviews*, vol. 21, no. 2, pp. 1985–1994, Feb. 2024, doi: 10.30574/wjarr.2024.21.2.0625.
- [32] A. V. Konstantinov and L. V. Utkin, “Interpretable machine learning with an ensemble of gradient boosting machines,” *Knowl. Based. Syst.*, vol. 222, Jun. 2021, doi: 10.1016/j.knosys.2021.106993.

- [33] B. Johansson and E. Ohlsson, "Gradient Boosting Machines and Nonlife Insurance Pricing Lecture Notes," 2022. [Online]. Available: <https://ssrn.com/abstract=4294965>
- [34] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, vol. 31, no. 6, pp. 3360–3379, 2023, doi: 10.1080/10494820.2021.1928235.
- [35] Y. Guang, "Generalized XGBoost Method," 2021.
- [36] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903–4923, Jul. 2024, doi: 10.1007/s10994-022-06296-4.
- [37] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," Jun. 15, 2024, *Elsevier Ltd.* doi: 10.1016/j.eswa.2023.122778.
- [38] N. A. Azhar, M. S. Mohd Pozi, A. M. Din, and A. Jatowt, "An investigation of SMOTE based methods for imbalanced datasets with data complexity analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 6651–6672, Jul. 2023, doi: 10.1109/TKDE.2022.3179381.
- [39] H. Hairani, T. Widiyaningtyas, and D. Dwi Prasetya, "INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION journal homepage : www.joiv.org/index.php/joiv INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION Addressing Class Imbalance of Health Data: a Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies." [Online]. Available: www.joiv.org/index.php/joiv
- [40] H. R. Sayegh, W. Dong, and A. M. Al-madani, "Enhanced Intrusion Detection with LSTM-Based Model, Feature Selection, and SMOTE for Imbalanced Data," *Applied Sciences (Switzerland)*, vol. 14, no. 2, Jan. 2024, doi: 10.3390/app14020479.
- [41] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, Sep. 2023, doi: 10.1109/TNNLS.2021.3136503.
- [42] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903–4923, Jul. 2024, doi: 10.1007/s10994-022-06296-4.
- [43] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
- [44] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.

- [45] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS One*, vol. 18, no. 10 October, Oct. 2023, doi: 10.1371/journal.pone.0291908.
- [46] Ş. K. Çorbacıoğlu and G. Aksel, "Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value," Oct. 01, 2023, *Wolters Kluwer Medknow Publications*. doi: 10.4103/tjem.tjem_182_23.
- [47] R. Yunis, "Hybridization Model for Air Pollution Prediction Using Time Series Data," *COGITO Smart Journal*, vol. 10, no. 1, 2024.
- [48] A. Sharbaf, "Feasibility Analysis of Data Science Methodologies in Architectural Design," Jun. 2025.
- [49] K. Shameti, "COMPARISON OF METHODOLOGICAL APPROACHES: CRISP-DM VERSUS OSEMN METHODOLOGY USING LINEAR REGRESSION AND STATISTICAL ANALYSIS," Jun. 2024.
- [50] S. E. Prastya, M. Adla, B. Nugraha, and Y. Sari, "Regression Analysis Data for E-Sport Athletes Prediction using OSEMN Framework," *Install: Information System and Technology Journal*, vol. 1, no. 1, pp. 1–8, Jun. 2024, doi: 10.33859/install.v1i1.542.
- [51] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, Feb. 2022, doi: 10.1007/s00607-021-00908-y.
- [52] X. Xiahou and Y. Harada, "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 2, pp. 458–475, Jun. 2022, doi: 10.3390/jtaer17020024.
- [53] Y. Liu, J. Fan, J. Zhang, X. Yin, and Z. Song, "Research on telecom customer churn prediction based on ensemble learning," *J. Intell. Inf. Syst.*, vol. 60, no. 3, pp. 759–775, Jun. 2023, doi: 10.1007/s10844-022-00739-z.
- [54] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system based on ensemble learning using feature grouping," *Applied Sciences (Switzerland)*, vol. 11, no. 11, Jun. 2021, doi: 10.3390/app11114742.
- [55] T. Gattermann-Itschert and U. W. Thonemann, "Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests," *Industrial Marketing Management*, vol. 107, pp. 134–147, Nov. 2022, doi: 10.1016/j.indmarman.2022.09.023.
- [56] K. Peng, Y. Peng, and W. Li, "Research on customer churn prediction and model interpretability analysis," *PLoS One*, vol. 18, no. 12 December, Dec. 2023, doi: 10.1371/journal.pone.0289724.
- [57] N. Ahmad, M. J. Awan, H. Nobanee, A. M. Zain, A. Naseem, and A. Mahmoud, "Customer Personality Analysis for Churn Prediction Using Hybrid Ensemble

- Models and Class Balancing Techniques,” *IEEE Access*, vol. 12, pp. 1865–1879, 2024, doi: 10.1109/ACCESS.2023.3334641.
- [58] M. Maw, S.-C. Haw, and C.-K. Ho, “Utilizing data sampling techniques on algorithmic fairness for customer churn prediction with data imbalance problems,” *F1000Res.*, vol. 10, p. 988, Sep. 2022, doi: 10.12688/f1000research.72929.1.
- [59] N Tahta Phudjashakty, Hasbi Firmansyah, Wahyu Asriyani, and Ali Sofyan, “Segmentasi Pelanggan Grosir Menggunakan K-Means: Analisis Outlier dan Ketidakseimbangan Data,” *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, vol. 4, no. 3, pp. 15993–16002, Jan. 2026, doi: 10.31004/jerkin.v4i3.4771.
- [60] I. Ekanayake and S. De Alwis, “Explainability, risk modeling, and segmentation based customer churn analytics for personalized retention in e-commerce,” Jan. 2026, doi: 10.1109/ICARC68737.2026.11454037.
- [61] W. Zhang *et al.*, “Integrating Network Pharmacology, Machine Learning, and Experimental Validation to Elucidate the Mechanism of Cardamonin in Treating Idiopathic Pulmonary Fibrosis,” *Int. J. Mol. Sci.*, vol. 27, no. 1, Jan. 2026, doi: 10.3390/ijms27010249.
- [62] M. Bogaert and L. Delaere, “Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art,” *Mathematics*, vol. 11, no. 5, Mar. 2023, doi: 10.3390/math11051137.

