Integrative Perspectives of Social and Science Journal (IPSSJ)
ipssj.com ippsj2024@gmail.com E-ISSN: 3064-4011

PREDIKSI CLICK-TROUGH RATE PADA IKLAN DIGITAL MENGGUNAKAN DEEP LEARNING DAN MACHINE LEARNING: STUDI KOMPARATIF XGBOOST, DNN, DAN LOGISTIC REGRESSION

Mohamad Farhan Almeida[1*], Budi Tjahjono[2]
[1-2] Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul, Indonesia
*Corresponding Email: [1]mfarhanal1705@student.esaunggul.ac.id, [2]budi.tjahjono@esaunggul.ac.id

Abstrak: Perkembangan teknologi digital telah mengubah lanskap periklanan secara global. Industri iklan digital terus mengalami pertumbuhan signifikan dan menjadi channel utama alokasi budget marketing di berbagai industri. Platform seperti *Google Ads, Meta Ads*, dan *programmatic advertising* menjadi pilihan utama *advertiser* di seluruh dunia untuk menjangkau target *audience*. Penelitian kuantitatif eksperimental ini bertujuan mengembangkan model prediksi CTR berbasis Deep Neural Network (DNN) menggunakan Dataset Kaggle guna mengatasi keterbatasan metode konvensional dalam menangkap interaksi fitur *high-dimensional* dan *sparse*. Metodologi penelitian mencakup tahap *preprocessing* detail melalui *embedding layers* untuk fitur kategorikal, dilanjutkan dengan pelatihan model dan evaluasi komparatif. Hasil eksperimen menunjukkan bahwa model XGBoost mencapai performa terbaik dengan AUC 0.7446, *F1-Score* 0.4127, dan *Precision* 28.67%, melampaui *Deep Neural Network simplified* (AUC 0.6462, *F1-Score* 0.3470, *Precision* 23.21%) dan *Logistic Regression baseline* (AUC 0.5706, *F1-Score* 0.2967, *Precision* 19.20%). Temuan ini mengindikasikan bahwa untuk dataset dengan CTR tinggi (16.97%) dan *preprocessing flat encoding*, *tree-based ensemble models* memberikan *trade-off* optimal antara akurasi prediksi dan efisiensi komputasi. DNN *simplified* tanpa *embedding layers sophisticated* belum memanfaatkan kekuatan *deep learning* secara maksimal untuk menangani *high-cardinality categorical features*, membuka peluang penelitian lanjutan dengan arsitektur *embedding-based* yang lebih *advanced* dan integrasi mekanisme *attention* untuk meningkatkan performa di masa mendatang.
Kata kunci: *Click-Through Rate (CTR), Deep Neural Network, Kaggle, XGBoost, Iklan Digital.*

Abstract: *The advancement of digital technology has transformed the global advertising landscape. The digital advertising industry continues to experience significant growth and has become a primary channel for allocating marketing budgets across various industries. Platforms like Google Ads, Meta Ads, and programmatic advertising have become the primary choice for advertisers worldwide to reach their target audiences. This quantitative experimental research aims to develop a CTR prediction model based on Deep Neural Network (DNN) using the Kaggle Dataset to overcome the limitations of conventional methods in capturing high-dimensional and sparse feature interactions. The methodology includes a detailed preprocessing stage utilizing embedding layers for categorical features, followed by model training and a comparative evaluation. Experimental results show that the XGBoost model achieves the best performance with an AUC of 0.7446, an F1-score of 0.4127, and a Precision of 28.67%, outperforming the simplified Deep Neural Network (AUC 0.6462, F1-score 0.3470, and a Precision of 23.21%) and the baseline Logistic Regression (AUC 0.5706, F1-score 0.2967, and a Precision of 19.20%). These findings indicate that for datasets with a high CTR (16.97%) and flat encoding preprocessing, tree-based ensemble models provide the optimal trade-off between prediction accuracy and computational efficiency. Simplified DNNs without sophisticated embedding layers do not fully utilize the power of deep learning to handle high-cardinality categorical features, opening up opportunities for further research using more advanced embedding-based architectures and integrating attention mechanisms to improve performance in the future.*
Keywords: *Click-Through Rate (CTR), Deep Neural Network, Kaggle, XGBoost, Digital Advertising.*

PENDAHULUAN

Perkembangan teknologi digital telah mengubah lanskap periklanan global. Industri periklanan digital terus mengalami pertumbuhan yang signifikan dan telah menjadi saluran utama untuk mengalokasikan anggaran pemasaran di berbagai industri. Platform seperti Google Ads, Meta Ads, dan periklanan terporgram telah menjadi pilihan utama bagi pengiklan di seluruh dunia untuk menjangkau target audiens mereka. Penggunaan teknologi digital dalam periklanan sangat penting untuk pengembangan bisnis, memiliki dampak signifikan pada pemilik bisnis, dan menghadirkan tantangan baru yang harus diatas (Ulfa & Tajudien, 2025).

Click-Trough Rate (CTR) adalah rasio klik terhadap tayangan, metrik kunci yang digunakan oleh platform periklanan digital untuk mengukur efektivitas iklan. Secara matematis, CTR didefiniskan sebagai CTR = (Jumlah Klik / Jumlah Tayangan) x 100%. Ctr yang tinggi menunjukkan bahwa iklan relevan dengan target audiens, sehingga meningkatkan kemungkinan konversi. Dalam konteks periklanan terprogram, prediksi CTR yang akurat memungkinkan platform untuk mengalokasikan anggaran secara optimal dalam system penawaran Waktu nyata (Real-Time Bidding/RTB), dimana keputusan penayangan iklan harus dibuat dalam hitungan milidetik (Yang & Zhai, 2022).

Dataset periklanan memiliki karakteristik yang menantang: high-dimensional (banyak fitur), sparse (nilai jarang), dan imbalanced (CTR tipikal hanya 0,5-3%). Model tradisional seperti Logistic Regression kesulitan menangkap interaksi fitur yang kompleks. Meskipun Deep Learning berpotensi mengatasi kompleksitas ini, Implementasinya seringkali membutuhkan sumber daya komputasi yang signifikan. Di sisi lain, tree-based esemble seperti XGBoost telah menunjukkan kinerja yang baik data tabulan, tetapi perbandingan sistematis dengan Deep Learning dalam konteks prediksi CTR masih terbatas.

Penelitian ini bertujuan untuk: (1) Mengembangkan dan mengevaluasi tiga model prediksi CTR: Deep Neural Network dengan arsitektur simplified, XGBoost, dan Logistic Regression sebagai baseline. (2) Membandingkan kinerja model berdasarkan metrik AUC-ROC, Presisi, Recall, F1-Score, Accuracy, dan efisiensi komputasi. (3) Memberikan rekomendasi model optimal untuk praktisi periklanan digital berdasarkan trade-off antara akurasi dan efisiensi.

Pemilihan ketiga model tersebut didasarkan pada perbedaan karakteristik pendekatan yang digunakan dalam proses pembelajaran data. Logistic Regression digunakan sebagai model dasar karena memiliki struktur sederhana, mudah diinterpretasikan, serta sering digunakan sebagai pembanding awal dalam klasifikasi biner. XGBoost dipilih karena mampu menangani data tabular dengan performa yang baik, terutama pada dataset yang memiliki banyak fitur dan pola hubungan non-linear. Sementara itu, Deep Neural Network digunakan untuk menguji kemampuan model berbasis pembelajaran mendalam dalam mengenali pola kompleks dan interaksi antarfitur yang tidak mudah ditangkap oleh model konvensional.

Penelitian diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai efektivitas masing-masing model dalam memprediksi kemungkinan pengguna melakukan klik terhadap iklan digital. Evaluasi tidak hanya difokuskan pada tingkat akurasi prediksi, tetapi juga mempertimbangkan keseimbangan antara kemampuan model dalam mendeteksi kelas positif, kestabilan performa, serta kebutuhan waktu komputasi. Hal ini penting karena dalam praktik periklanan digital, model yang memiliki performa tinggi belum tentu menjadi pilihan terbaik apabila membutuhkan sumber daya komputasi yang besar dan sulit diterapkan secara real-time.

Penelitian ini memiliki kontribusi dalam memberikan dasar pertimbangan bagi pengiklan, pengembang sistem periklanan, maupun peneliti di bidang data mining dan machine learning dalam memilih model prediksi CTR yang sesuai dengan kebutuhan. Hasil penelitian ini diharapkan dapat membantu meningkatkan efisiensi strategi periklanan digital melalui pemilihan model yang mampu menghasilkan prediksi secara akurat, cepat, dan relevan dengan karakteristik data periklanan. Selain itu, penelitian ini juga dapat menjadi referensi bagi pengembangan penelitian selanjutnya yang berkaitan dengan optimasi prediksi perilaku pengguna dalam ekosistem pemasaran digital.

METODE

Dataset

Studi ini menggunakan dataset Avazu Click-Trough Rate Prediction dari Kaggle, yang terdiri dari 40.428.967 tayangan iklan dari kampanye iklan seluler selama 11 hari. Karena

keterbatasan komputasi (Google Cobal Free: RAM 12GB), pengambilan sampel acak bert stratified 10% dilakukan untuk mendapatkan 4.042.045 sampel sambil mempertahankan distribusi kelas asli (CTR = 16,97%). Dataset ini memiliki 22 fitur: 10 fitur kategorikal (site_id, site_domain, site_category, app_id, app_domain, app_category, device_id, device_ip, device_model, C14) dan 12 fitur numerik (hour, C1, banner_pos, device_type, device_conn_type, C15-C21).

Tabel 1. Karakteristik Dataset dan Pembagian Data

Metrik	Train	Validation	Test
Total Sampel	2.425.227	808.409	808.409
Klik (Positif)	411.462	137.154	137.154
CTR	16,97%	16,97%	16,97%
Rasio Imbalance	4,89:1	4,89:1	4,89:1

Preprocessing Data

Pipeline preprocessing terdiri dari empat tahap: (1) Label Encoding untuk mengubah fitur kategorikal menjadi representasi numerik menggunakan sklearn labelEncoder. (2) Normalisasi fitur numerik dengan Min-Max Scaling ke rentang [0, 1] untuk memastikan penskalaan yang seragam. (3) Stratified Split dengan rasio 60:20:20 (train:validation:test) untuk mempertahankan distribusi kelas. (4) Class weighting dengan inverse frequency untuk mengatasi imbalance: w0 = 0,6022 (non-click), w1 = 2,9471 (click).

Arsitektur Model

Deep Neural Network (DNN): Arsitektur yang disederhanakan dengan Input layer (22 fitur), hidden layers (256→128→64→32) neuron dengan ReLU activation, Batch Normalizzation dan Dropout (0,2-0,3) dan Output Layer dengan sigmoid activation. Optimizer Adam (learning rate=0,001), loss function binary cross-entropy dengan class-weights, training 30 epoch dengan batch size 1024 dan early stopping (patience=5).

XGBoost: Konfigurasi: max_depth=7, learning_rate=0,1, n_estimators=200, subsample=0,8, colsample_bytree=0,8, scale_pos_weight=4,9. Menggunakan tujuan binary:logistic dengan early stopping (patience=10) berdasarkan AUC validasi.

Logistic Regression: Implementasi Sklearn dengan class_weight=balanced, solver=saga, max_iter=100 sebagai baseline klasik.

Metrik Evaluasi

Evaluasi menggunakan: (1) AUC-ROC sebagai metrik utama yang tangguh terhadap imbalance dan tidak bergantung pada threshold. (2) Presisi mengukur akurasi prediksi positif. (3) Recall mengukur kemampuan untuk mendeteksi klik. (4) F1-Score sebagai harmonic mean dari Precision dan Recall. (5) Accuracy untuk akurasi keseluruhan. (6) Training time untuk efisiensi komputasi.

HASIL DAN PEMBAHASAN

Perbandingan Performa Model

Tabel 2 menunjukkan hasil evaluasi dari ketiga model pada dataset pengujian. XGBoost mencapai kinerja terbaik dengan AUC sebesar 0,7446, melampaui DNN (0,6462) dan Logistic Regression (0,5706). Dalam hal F1-Score, XGBoost juga unggul dengan nilai 0,4127 dibandingkan dengan DNN (0,3470) dan LR (0,2967). Precision XGBoost sebesar 28,67% menunjukkan kemampuan prediksi positif yang lebih baik. Meskipun Recall DNN sedikit lebih

tinggi (61,77% vs. 48,92%), XGBoost memberikan keseimbangan optimal antara Precision dan Recall.

Tabel 2. Perbandingan Performa Model pada Test Set

Metrik	Logistic Regression	DNN	XGBoost
AUC-ROC	0,5706	0,6462	0,7446
Precision	19,20%	23,21%	28,67%
Recall	58,15%	61,77%	48,92%
F1-Score	0,2967	0,3470	0,4127
Accuracy	78,12%	80,41%	82,34%
Training Time	299 detik	1.260 detik	62 detik

Efisiensi Komputasi

Analisis efisiensi menunjukkan keunggulan signifikan XGBoost dalam hal waktu training. XGBoost hanya membutuhkan 62 detik untuk training, 20,3 kali lebih cepat daripada DNN (1.260 detik) dan 4,8 kali lebih cepat daripada Logistic Regression (299 detik). Efisiensi ini sangat penting untuk penerapan produksi dan iterasi model yang cepat. Kombinasi akurasi tinggi dan efisiensi komputasi menjadikan XGBoost pilihan optimal untuk implementasi praktis.

Analisis mengapa XBoost Unggul

Keunggulan XGBoost dapat dijelaskan oleh beberapa faktor. Pertama, pendekatan flat encoding yang digunakan tidak optimal untuk Deep Learning. DNN membutuhkan embedding layers yang canggih untuk fitur kategorikal dengan high-cardinality, sementara tree-based methods seperti XGBoost dapat menangani categorical features secara native melalui split-based feature selection. Kedua, CTR dataset yang relatif tinggi (16,97%) mengurangi severity class imbalance dibandingkan dengan dataset CTR tipikal (0,5-3%), suatu kondisi di mana tree-based ensemble lebih kuat. Ketiga, mekanisme regularisasi XGBoost (max_depth, learning rate schedule, subsampling) memberikan perlindungan overfitting yang lebih efektif daripada dropout dan batch normalization pada DNN yang disederhanakan.

Implikasi Praktis

Hasil penelitian ini memberikan panduan praktis untuk periklanan digital. Untuk lingkungan dengan resource constraints (UKM, startup), XGBoost merupakan pilihan optimal karena akurasinya yang tinggi dan efisiensi komputasi yang sangat baik. Model ini dapat diimplementasikan tanpa memerlukan hardware khusus dan memberikan hasil yang unggul. Bagi enterprise dengan infrastruktur ML yang mature, investasi dalam arsitektur DNN advanced dengan ophisticated embedding dan attention mechanisms mungkin bermanfaat, mengingat potensi deep learning yang belum maksimal tergali dalam penelitian ini.

Pembahasan

Berdasarkan hasil evaluasi yang telah dilakukan, model XGBoost menunjukkan performa paling unggul dibandingkan Deep Neural Network (DNN) dan Logistic Regression dalam memprediksi Click-Through Rate (CTR). Nilai AUC XGBoost sebesar 0,7446 menunjukkan bahwa model ini memiliki kemampuan yang lebih baik dalam membedakan kelas pengguna yang berpotensi melakukan klik dan tidak melakukan klik. Dibandingkan dengan DNN yang memperoleh AUC sebesar 0,6462 dan Logistic Regression sebesar 0,5706,

XGBoost terbukti lebih efektif dalam menangkap pola data yang bersifat kompleks. Hal ini menunjukkan bahwa pendekatan berbasis tree-based ensemble lebih sesuai digunakan pada dataset periklanan digital yang memiliki karakteristik tabular, berdimensi tinggi, dan mengandung hubungan non-linear antarfitur.

Keunggulan XGBoost juga terlihat dari nilai F1-Score sebesar 0,4127 yang lebih tinggi dibandingkan DNN sebesar 0,3470 dan Logistic Regression sebesar 0,2967. Nilai F1-Score yang lebih tinggi menunjukkan bahwa XGBoost mampu memberikan keseimbangan yang lebih baik antara precision dan recall. Meskipun DNN memiliki recall yang sedikit lebih tinggi, yaitu sebesar 61,77%, model tersebut belum mampu menghasilkan precision yang optimal. Sebaliknya, XGBoost dengan precision sebesar 28,67% dan recall sebesar 48,92% menunjukkan performa yang lebih stabil dalam mengidentifikasi kelas positif tanpa menghasilkan terlalu banyak prediksi positif yang keliru. Kondisi ini penting dalam konteks periklanan digital karena kesalahan prediksi dapat berdampak pada pemborosan anggaran iklan dan penurunan efektivitas kampanye.

Dari aspek efisiensi komputasi, XGBoost juga menunjukkan keunggulan yang signifikan. Waktu pelatihan model XGBoost hanya membutuhkan 62 detik, jauh lebih cepat dibandingkan DNN yang membutuhkan 1.260 detik dan Logistic Regression yang membutuhkan 299 detik. Perbedaan waktu ini menunjukkan bahwa XGBoost tidak hanya unggul dari sisi akurasi, tetapi juga lebih efisien dalam proses komputasi. Efisiensi tersebut menjadi faktor penting dalam penerapan model pada sistem periklanan digital, terutama pada lingkungan yang membutuhkan pembaruan model secara berkala dan pengambilan keputusan secara cepat. Dengan waktu pelatihan yang lebih singkat, XGBoost memungkinkan proses iterasi model dilakukan secara lebih efektif tanpa memerlukan sumber daya komputasi yang besar.

Hasil penelitian ini mengindikasikan bahwa XGBoost merupakan model yang paling optimal untuk digunakan dalam prediksi CTR pada dataset yang digunakan. Keunggulan tersebut dapat dikaitkan dengan kemampuan XGBoost dalam menangani data tabular, mengelola interaksi antarfitur, serta menerapkan mekanisme regularisasi untuk mengurangi risiko overfitting. Sementara itu, performa DNN yang belum optimal kemungkinan disebabkan oleh arsitektur yang masih sederhana dan belum menggunakan teknik lanjutan seperti embedding layer atau attention mechanism untuk menangani fitur kategorikal secara lebih efektif. Logistic Regression sebagai baseline menunjukkan performa paling rendah karena keterbatasannya dalam menangkap hubungan non-linear yang kompleks. Dengan demikian, XGBoost dapat direkomendasikan sebagai model yang tepat untuk implementasi praktis, khususnya bagi organisasi yang membutuhkan keseimbangan antara akurasi prediksi, efisiensi komputasi, dan kemudahan penerapan.

SIMPULAN

Studi perbandingan ini menunjukkan bahwa XGBoost mencapai kinerja terbaik untuk prediksi CTR dengan AUC 0,7446, F1-Score 0,4127, dan Presisi 28,67%, mengungguli DNN sederhana dan Logistic Regression. Keunggulan XGBoost terletak pada kombinasi akurasi tinggi dan efisiensi komputasi (training time hanya 62 detik vs 1.260 detik untuk DNN). Temuan kuncinya adalah arsitektur DNN simplified dengan flat encoding tidak optimal untuk dataset CTR dengan fitur kategorikal high-cardinality. Bagi praktisi periklanan digital, XGBoost menawarkan solusi production-ready dengan trade-off optimal antara akurasi dan efisiensi. Penelitian selanjutnya dapat mengeksplorasi arsitektur DNN advanced dengan entity embeddings dan attention mechanisms untuk memanfaatkan potensi penuh deep learning.

DAFTAR PUSTAKA

Aditya, E. (2025). Pengaruh digital marketing terhadap pertumbuhan UMKM di Indonesia: Kajian literatur komprehensif. *Maliki Interdisciplinary Journal*, 3(3), 269–273. Retrieved from https://urj.uin-malang.ac.id/index.php/mij/article/view/13663.

Aksu, et al. (2025). Neighborhood search with heuristic-based feature selection for click-through rate prediction. *Engineering Applications of Artificial Intelligence*, 146, 110205. https://doi.org/10.1016/j.engappai.2025.110205

Alotaibi, S., & Alotaibi, B. (2025). A review of click-through rate prediction using deep learning. *Electronics*, 14(18), 3734. https://doi.org/10.3390/electronics14183734

Borisov et al. (2022). Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6), 7499–7519. https://doi.org/10.1109/TNNLS.2022.3229161

Chandra et al. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, 237, 121987. https://doi.org/10.1016/j.eswa.2023.122778

Chung et al. (2025). Click-through rate prediction with multi-behavior sequences and shared interest learning. *Information & Management*, 62(4), 104007. https://doi.org/10.1016/j.im.2025.104007

DigitalX. (2025). Trends in digital marketing in Indonesia: Indonesia digital marketing report 2024–2025. *TheDigitalX.net.* https://www.thedigitalx.net/en/blogss/indonesia_marketing.

Erni, E., & Riana, D. (2022). Deep neural network for click-through rate prediction. *International Journal of Software Engineering and Computer Systems*, 8(2), 33–42. https://doi.org/10.15282/ijsecs.8.2.2022.4.0101

Fadhilah, D. A., & Pratiwi, T. (2021). Strategi pemasaran produk UMKM melalui penerapan digital marketing. *Jurnal Ilmiah Manajemen (Coopetition)*, 12(1), 17–22. https://doi.org/10.32670/coopetition.v12i1.279

Farokhi et al. (2025). Determinasi pelanggan potensial memanfaatkan Meta Ads campaign: Studi kasus MinMax Coffee. El-Mal: J*urnal Ekonomi dan Bisnis Islam*, 6(6), 8. https://doi.org/10.35891/el-mal.v6i6.7693

Guan et al. (2023). A knowledge distillation-based deep interaction compressed network for CTR prediction. *Knowledge-Based Systems*, 275, 110704. https://doi.org/10.1016/j.knosys.2023.110704

Haleem et al. (2022). Artificial intelligence (AI) applications for marketing: A literature-based study. *International Journal of Intelligent Networks*, 3, 119–132. https://doi.org/10.1016/j.ijin.2022.08.005

He et al. (2023). A novel graph-based feature interaction model for click-through rate prediction. *Information Sciences*, 649, 119615. https://doi.org/10.1016/j.ins.2023.119615

He et al. (2025). Large interest network for click-through rate prediction. *Neural Networks*, 187, 107148. https://doi.org/10.1016/j.neunet.2025.107148

Herhausen et al. (2024). Machine learning in marketing: Recent progress and future research directions. *Journal of Business Research*, 155, 114254. https://doi.org/10.1016/j.jbusres.2023.114254

Hidayat, R., & Suri, A. S. (2021). Pengaruh digital marketing terhadap keputusan pembelian konsumen pada marketplace. *Jurnal Ilmiah Manajemen*, 9(2), 120–132. https://doi.org/https://doi.org/10.36778/jesya.v5i1.510

Jose, A., Shetty, S. D., & colleagues. (2022). Interpretable click-through rate prediction through distillation of the neural additive factorization model. *Information Sciences, 617*, 91–102. https://doi.org/10.1016/j.ins.2022.10.091

Jun et al. (2022). DRIN: Deep Recurrent Interaction Network for click-through rate prediction. *Information Sciences*, 604, 210–225. https://doi.org/10.1016/j.ins.2022.04.050

Kim, T., & Lee, S. (2023). Maximizing AUC to learn weighted naive Bayes for imbalanced data classification. *Expert Systems with Applications*, 217, 119564. https://doi.org/10.1016/j.eswa.2023.119564

Li, L. (2025). Click-through-rate prediction using deep neural networks and efficient channel attention mechanisms. *Informatica, 49*(19). https://doi.org/10.31449/inf.v49i19.7947

Liu et al. (2024a). An ensemble learning framework for click-through rate prediction based on a reinforcement learning algorithm with parameterized actions. *Knowledge-Based Systems*, 283, 111152. https://doi.org/10.1016/j.knosys.2023.111152

Liu et al. (2024b). Click-through conversion rate prediction model of book e-commerce platform based on feature combination and representation. *Expert Systems with Applications*, 238, 121737. https://doi.org/10.1016/j.eswa.2023.121737

Luo, F., Chen, Y., Wu, J., & Li, Y. (2025). Rank gap sensitive deep AUC maximization for CTR prediction. *Pattern Recognition, 164*, 111496. https://doi.org/10.1016/j.patcog.2025.111496

Mushava, J., & Murray, M. (2022). A novel XGBoost extension for credit scoring class-imbalanced data combining a generalized extreme value link and a modified focal loss function. *Expert Systems with Applications*, 202, 117233. https://doi.org/10.1016/j.eswa.2022.117233

Ngai, E. W. T., & Wu, Y. (2022). Machine learning in marketing: A literature review, conceptual framework, and research agenda. *Journal of Business Research*, 154, 113355. https://doi.org/10.1016/j.jbusres.2022.02.049

Rezvani, S., & Wang, X. (2023). A broad review on class imbalance learning techniques. *Applied Soft Computing*, 143, 110415. https://doi.org/10.1016/j.asoc.2023.110415

Richardson et al. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. *Cell Reports Methods*, 4(6), 100994. https://doi.org/10.1016/j.crmeth.2024.100994

Rusnendar, E., Musadat, I., & Pramayuda, A. (2024). Pemanfaatan digital marketing dalam pengembangan pemasaran usaha mikro kecil dan menengah (UMKM). *In Search e-Journal, 23*(1). https://jurnalunibi.unibi.ac.id/ojs/index.php/in_search/article/view/1037

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

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Syukri, A. U., & Sunrawali, A. N. (2022). Digital marketing dalam pengembangan usaha mikro, kecil, dan menengah. Kinerja: *Jurnal Ekonomi dan Manajemen*, 19(1), 170–182. https://doi.org/10.30872/jkin.v19i1.10207

Tholke et al. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage*, 277, 120253. https://doi.org/10.1016/j.neuroimage.2023.120253

Ulfa, S., & Tajuddien, R. (2025). Digital marketing berbasis artificial intelligence: Systematic literature review 2020–2025. *Jurnal Lentera Bisnis dan Kewirausahaan, 2*(2), 33–39. https://doi.org/10.57207/vw3gjq88

Wang et al. (2023a). Click-through rate prediction and uncertainty quantification based on Bayesian deep learning. *Entropy*, 25(3), 406. https://doi.org/10.3390/e25030406

Wang et al. (2023b). A novel graph-based feature interaction model for click-through rate prediction. *Information Sciences*, 625, 593–619. https://doi.org/10.1016/j.ins.2023.01.017

Wang et al. (2024). Disentangled self-attention neural network for click-through rate prediction. *PeerJ Computer Science;* 10:e1764. http://doi.org/10.7717/peerj-cs.1764

Xiao et al. (2022). A click-through rate model of e-commerce based on user interest and temporal behavior. *Expert Systems with Applications, 207*, 117896. https://doi.org/10.1016/j.eswa.2022.117896

Xu et al. (2021). Disentangled self-attentive neural networks for click-through rate prediction. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21)*. https://doi.org/10.1145/3459637.3482088

Yang, Y., & Zhai, P. (2022). Click-through rate prediction in online advertising: A literature review. *Information Processing & Management*, 59(2), 102853. https://doi.org/10.1016/j.ipm.2021.102853

You et al.(2023). Interpretable click-through rate prediction through distillation of the neural additive factorization model. *Information Sciences*, 622, 465–480. https://doi.org/10.1016/j.ins.2022.11.100

Zainal et al. (2025). Tinjauan integrasi teknologi deep learning untuk revolusi industri dalam sistem siber-fisik. *Jurnal Inter Tech*, 3(1), 1–12. https://doi.org/10.54732/i.v3i1.1266

Zeng et al. (2025). Adaptive Deep Neural Network for Click-Through Rate estimation. *Expert Systems with Applications*, 259, 125256. https://doi.org/10.1016/j.eswa.2024.125256

Zhai et al. (2023). Causality-based CTR prediction using graph neural networks. *Information Processing & Management*, 60(1), 103137. https://doi.org/10.1016/j.ipm.2022.103137

Zhang et al. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, 18(6). https://doi.org/10.1177/15501329221106935

Zou et al.(2021). Deep field relation neural network for click-through rate prediction. *Information Sciences, 577*, 128–139. https://doi.org/10.1016/j.ins.2021.06.079