

## OPTIMASI PREDIKSI RISIKO KREDIT DENGAN PREPROCESSING DAN HYPERPARAMETER TUNING

Amin Nur Rais<sup>1)</sup>, Warjiyono<sup>2)</sup>, Jordy Lasmana Putra<sup>3)</sup>

<sup>1</sup> Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika  
Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta

<sup>2</sup> Program Studi Sistem Informasi Akuntansi Kampus Kota Tegal, Fakultas Teknik dan Informatika, Universitas BSI  
Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta

<sup>3</sup> Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri  
Jl. Raya Jatiwaringin No.2, RT.8/RW.13, Cipinang Melayu, Kec. Makasar, Kota Jakarta Timur, DKI Jakarta, Indonesia  
Co Responden Email: amin.arv@bsi.ac.id

### Abstract

#### Article history

Received 11 Okt 2024

Revised 12 Dec 2024

Accepted 31 Dec 2024

Available online 31 Jan 2025

#### Keywords

Credit Risk,  
Machine Learning,  
Gradient Boosting,  
KNN

*Credit risk is a challenge in the financial industry, which can impact the stability of financial institutions. This study evaluates the performance of machine learning models in predicting credit risk using a dataset from Kaggle. Four models were tested: Logistic Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbors (KNN), each tested in three versions: baseline, preprocessing, and tuned. The preprocessing process includes handling missing values, encoding categorical features, and standardizing numerical features. Models were evaluated based on accuracy, precision, recall, and F1-score. The results show that the Gradient Boosting (Tuned) model performs the best with an accuracy of 93.79%, precision of 94.91%, recall of 76.05%, and F1-score of 84.44%. This research provides benefits for financial institutions in selecting the optimal model for predicting credit risk and supporting data-driven decision-making.*

### Abstrak

#### Riwayat

Diterima 11 Okt 2024

Revisi 11 Nov 2024

Disetujui 31 Des 2024

Terbit online 31 Jan 2025

#### Kata Kunci

Risiko Kredit,  
Pembelajaran Mesin,  
Gradient Boosting,  
KNN

Risiko kredit menjadi tantangan dalam industri keuangan, yang dapat berdampak pada stabilitas lembaga keuangan. Penelitian ini mengevaluasi kinerja model machine learning dalam memprediksi risiko kredit menggunakan dataset dari Kaggle. Empat model yang diuji adalah Logistic Regression, Random Forest, Gradient Boosting, dan K-Nearest Neighbors (KNN), yang masing-masing diuji dalam tiga versi: baseline, preprocessing, dan tuned. Proses preprocessing mencakup penanganan nilai hilang, encoding fitur kategori, dan standarisasi fitur numerik. Model dievaluasi berdasarkan akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Gradient Boosting (Tuned) memberikan performa terbaik dengan akurasi 93.79%, presisi 94.91%, recall 76.05%, dan F1-score 84.44%. Penelitian ini memberikan manfaat bagi lembaga keuangan dalam memilih model yang optimal untuk memprediksi risiko kredit dan mendukung pengambilan keputusan berbasis data.

## PENDAHULUAN

Penilaian risiko kredit menjadi elemen yang penting dalam industri perbankan dan keuangan, karena menentukan kemampuan lembaga keuangan untuk mengelola portofolio kredit secara efektif dan menjaga stabilitas finansial lembaga. Risiko kredit mengacu pada kemungkinan gagal bayar oleh peminjam, yang jika tidak dikelola dengan baik dapat merugikan stabilitas operasional bank (Bello 2023). Seiring berjalannya waktu, data

keuangan semakin kompleks, sehingga metode konvensional yang digunakan dalam analisis risiko kredit sering kali terbatas untuk menangkap pola-pola yang kompleks dan dinamis, sehingga diperlukan pendekatan yang lebih canggih dan akurat untuk menyelesaikan permasalahan tersebut seperti menggunakan pembelajaran mesin (*machine learning*) (Wang et al. 2023).

*Machine learning* telah menjadi alat yang populer dalam melakukan penilaian risiko kredit karena kemampuannya untuk

memproses data dalam skala besar, seperti data nasabah dan informasi keuangan, dan mendeteksi pola yang tidak terlihat oleh metode statistik tradisional (Kurniawan et al. 2022). Teknik *machine learning* seperti algoritma *decision tree* dan *ensemble* dapat meningkatkan prediksi kualitas kredit dalam situasi data besar dan seimbang dengan hasil akurasi yang lebih baik (Normah et al. 2022). Penerapan metode ini memberikan keunggulan dalam menilai berbagai faktor yang berkontribusi pada risiko kredit, mulai dari riwayat kredit hingga faktor ekonomi makro yang mempengaruhi kelayakan peminjam (P, S, dan Jaiswal 2024).

Di Indonesia, penerapan pembelajaran mesin dalam penilaian kualitas kredit baru saja mulai berkembang, dengan adanya beberapa penelitian menunjukkan bahwa model seperti *Random Forest* dan *Deep Neural Network* dapat menghasilkan akurasi tinggi dalam prediksi pinjaman yang mungkin bermasalah (Kurniawan et al. 2022)(Normah et al. 2022).

Penelitian ini bertujuan untuk menentukan algoritma yang paling efektif dalam melakukan prediksi risiko kredit menggunakan teknik *machine learning* dengan membandingkan beberapa model klasifikasi, yaitu *Logistic Regression*, *Random Forest*, *Gradient Boosting*, dan *K-Nearest Neighbors*. Berdasarkan hasil evaluasi pada metrik seperti akurasi, presisi, recall, dan F1-score, penelitian ini akan mengidentifikasi model yang dapat memberikan prediksi yang akurat dan efisien pada data kredit dengan mempertimbangkan efek dari preprocessing dan tuning parameter. Hasil perbandingan ini akan memberikan panduan bagi lembaga keuangan dalam memilih model yang optimal untuk memitigasi risiko kredit serta meningkatkan proses pengambilan keputusan berbasis data.

## METODE PENELITIAN

Penelitian ini dilakukan untuk meningkatkan evaluasi prediksi risiko kredit dengan lima tahap seperti yang ditunjukkan pada Gambar 1 :

### 1. Pengumpulan Data

Pada tahap pengumpulan data pada penelitian ini menggunakan dataset publik yang bersumber dari *kaggle repository* dengan nama dataset *Credit Risk Dataset*. *Dataset* tersebut terdiri dari 32.581 data yang berisikan informasi usia, pendapatan

tahunan, status kepemilikan rumah, lama bekerja, tujuan pinjaman, grade pinjaman, jumlah pinjaman, tingkat bunga, status pinjaman (gagal bayar atau tidak), persentase pinjaman terhadap pendapatan, riwayat gagal bayar, dan lama riwayat kredit.

### 2. Pra Pemrosesan Data

Pada tahap pra pemrosesan data dilakukan untuk memastikan konsistensi dan kualitas data sebelum dilakukan tahap pemodelan. Langkah-langkah yang dilakukan meliputi penanganan nilai hilang pada fitur numerik, perubahan fitur kategori menjadi nilai numerik, dan standarisasi fitur numerik untuk menyamakan skala. Jumlah data yang digunakan dalam tahap pra pemrosesan sebanyak 32.581 data yang diambil dari *dataset* yang tersedia. Pengambilan jumlah data sebanyak ini dengan mempertimbangkan semakin banyak data yang digunakan, semakin baik model dapat menangkap pola-pola yang ada, sehingga meningkatkan kemampuan generalisasi, mengurangi potensi bias dan meningkatkan kompatibilitas data dengan model *machine learning* yang akan digunakan (Fan et al. 2021).

### 3. Pembagian Data

Pada tahap pembagian data, data dibagi secara acak menjadi data pelatihan dan data pengujian dengan rasio 80:20. Tahap pembagian data dilakukan untuk memastikan bahwa model memiliki data yang cukup untuk proses belajar (*training*) sekaligus data yang mencukupi untuk proses uji performa model secara independen (*testing*). Pembagian data dilakukan secara acak untuk mengurangi potensi bias (Muningsih 2022).

### 4. Pemodelan

Pada tahap pemodelan dilakukan dengan menggunakan 4 model *machine learning* :

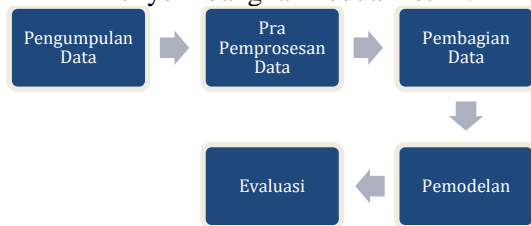
- a) Logistic Regression.
- b) Random Forest.
- c) Gradient Boosting.
- d) K-Nearest Neighbors (KNN)

Dari ke empat model *machine learning* akan diujikan kedalam 3 model pengujian: model baseline tanpa *preprocessing* dan *tuned*, model dengan *preprocessing*, dan model *tuned* dengan *hyperparameter tuning*.

## 5. Evaluasi

Pada tahap evaluasi digunakan untuk mengukur performa model pada setiap pemodelan yang dilakukan menggunakan empat metrik evaluasi:

- Akurasi: Mengukur proporsi prediksi yang benar dari seluruh prediksi.
- Presisi: Mengukur ketepatan prediksi positif dibandingkan seluruh prediksi positif.
- Recall: Mengukur sensitivitas model dalam mendeteksi kelas positif.
- F1-Score: Rata-rata harmonis dari presisi dan recall untuk menyeimbangkan kedua metrik.



Gambar 1. Tahapan Penelitian

## HASIL DAN PEMBAHASAN

Penelitian ini menggunakan aplikasi berbasis website yaitu Google Colab untuk melakukan tahapan penelitian sebagai alat simulasinya (Gelar Guntara 2023).

### A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini bersumber dari data repositori kaggle.com dengan nama "Credit Risk Dataset" oleh Laotse, yang memiliki data sebanyak 32.581 data dengan 11 fitur dan satu label kelas biner, di mana kelas "0" menunjukkan tidak gagal bayar dan kelas "1" menunjukkan gagal bayar. Dari 32.581 data, terbagi menjadi 25.473 data kelas 0, dan 7.108 dari kelas 1, menunjukkan ketidakseimbangan kelas yang sering ditemukan dalam data risiko kredit. Dataset ini akan digunakan untuk pelatihan, validasi, dan pengujian model dalam upaya mengidentifikasi model terbaik untuk memprediksi risiko kredit, dengan mempertimbangkan tahap *preprocessing* dan *hyperparameter tuning*.

Tabel 1. Deskripsi Dataset

Nama Fitur	Deskripsi
person_age	Usia
person_income	Pendapatan Tahunan

person_home_ownership	Status Kepemilikan Rumah
person_emp_length	Lama Bekerja (dalam tahun)
loan_intent	Tujuan Pinjaman
loan_grade	Tingkat Pinjaman
loan_amnt	Jumlah Pinjaman (dalam dolar)
loan_int_rate	Suku Bunga
loan_percent_income	Persentase Pendapatan
cb_person_default_on_file	Riwayat Gagal Bayar Lama Riwayat
cb_person_cred_hist_length	Kredit (dalam tahun)
loan_status	Status Pinjaman

### B. Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan melalui beberapa langkah. Pertama, nilai hilang pada fitur numerik seperti lama bekerja dan suku bunga pinjaman diisi dengan median. Kedua, fitur kategori seperti kepemilikan rumah dan tujuan pinjaman dikonversi ke nilai numerik menggunakan Label Encoding. Terakhir, fitur numerik seperti usia dan pendapatan distandarisasi dengan Standard Scaler untuk menyamakan skala, sehingga meningkatkan kinerja model machine learning. Tahapan ini bertujuan untuk memastikan data yang digunakan konsisten dan siap untuk analisis lebih lanjut.

```

# Handle missing values
imputer = SimpleImputer(strategy="median")
data[["person_emp_length", "loan_int_rate"]] = imputer.fit_transform(data[["person_emp_length", "loan_int_rate"]])

# Encode categorical features
categorical_columns = ["person_home_ownership", "loan_intent", "loan_grade", "cb_person_default_on_file"]
for column in categorical_columns:
    le = LabelEncoder()
    data[column] = le.fit_transform(data[column])

# Standardize numerical features
numerical_columns = ["person_age", "person_income", "person_emp_length", "loan_amnt", "loan_int_rate", "loan_percent_income", "cb_person_cred_hist_length"]
scaler = StandardScaler()
data[numerical_columns] = scaler.fit_transform(data[numerical_columns])

# Separate features and target variable
X = data.drop("loan_status", axis=1)
y = data["loan_status"]
  
```

Gambar 2. Proses Preprocessing

### C. Pembagian Data

Pembagian data dalam machine learning dapat dilakukan dengan rasio 80:20 antara data pelatihan dan data pengujian. Rasio ini memberikan cukup data pada proses pelatihan untuk memungkinkan model

belajar secara mendalam dari pola dalam data, yang pada tahap selanjutnya dapat meningkatkan akurasi dan prediksi. Sementara itu, data pengujian sebesar 20% memberikan ukuran yang representatif dan cukup untuk menilai performa model pada data yang belum pernah dilihat sebelumnya, sehingga memperkuat kemampuan model (Raschka 2018). Studi menunjukkan bahwa penggunaan rasio ini membantu mengurangi risiko overfitting, memastikan bahwa model yang dilatih tidak hanya baik pada data pelatihan tetapi juga pada data baru (Roy et al. 2019).

#### D. Pemodelan

Pemodelan dilakukan dengan empat algoritma utama: Logistic Regression, Random Forest, Gradient Boosting, dan K-Nearest Neighbors (KNN) (Budiman et al. 2023). Setiap model diuji dalam tiga tahap:

##### 1. Baseline

Model pertama dilatih dan diuji tanpa preprocessing dan hyperparameter tuning untuk mendapatkan metrik baseline, seperti akurasi, presisi, recall, dan F1-score (Shem L. Gonzales 2023). Hasil ini berfungsi sebagai referensi awal untuk perbandingan. Pada tabel 2 diketahui bahwa hasil pengujian baseline menunjukkan bahwa Random Forest memiliki performa terbaik dengan akurasi 93.0%, precision 96.4%, recall 71.1%, dan F1-score 81.8%, mengindikasikan keseimbangan yang kuat antara precision dan recall. Gradient Boosting mengikuti dengan akurasi 92.4% dan F1-score 80.1%, sementara Logistic Regression hanya mencapai recall 44%, menghasilkan F1-score 54.5%, dan K-Nearest Neighbors (KNN) memiliki performa lebih rendah dengan akurasi 89.2% dan F1-score 71.9%. Secara keseluruhan, Random Forest menonjol sebagai model baseline terbaik dan menjadi kandidat utama untuk peningkatan lebih lanjut.

Tabel 2. Hasil Pengujian Baseline

Model	TN	FP	FN	TP	Accuracy	Precision	Recall	F1-Score
Logistic Regression (Baseline)	4,820	252	809	636	0.837	0.716	0.440	0.545
Random Forest (Baseline)	5,034	38	418	1,027	0.930	0.964	0.711	0.818
Gradient Boosting (Baseline)	5,015	57	441	1,004	0.924	0.946	0.695	0.801
K-Nearest Neighbors (Baseline)	4,906	166	541	904	0.892	0.845	0.626	0.719

##### 2. Preprocessing

Model kedua, model diuji dengan data yang telah melalui proses preprocessing (Kulsum, Jajuli, dan Sulistiyowati 2022). Tahap ini memungkinkan analisis pengaruh preprocessing terhadap kinerja model. Pada tabel 3 diketahui bahwa hasil pengujian model kedua menunjukkan bahwa kinerja setiap model tidak mengalami perubahan signifikan dibandingkan dengan baseline. Logistic Regression tetap memiliki akurasi 83.7%, precision 71.6%, recall 44%, dan F1-score 54.5%, menunjukkan bahwa preprocessing tidak meningkatkan kemampuan deteksi positif model ini. Random Forest juga menunjukkan stabilitas dengan akurasi 93.0%, precision 96.4%, recall 71.1%, dan F1-score 81.8%, tetap menjadi model terbaik dalam hal keseimbangan metrik. Gradient Boosting mempertahankan akurasi 92.4% dengan F1-score 80.1%, sementara K-Nearest Neighbors (KNN) memiliki akurasi 89.2% dan F1-score 71.9%. Dengan demikian, preprocessing tidak memberikan perubahan substansial

pada performa model, tetapi Random Forest tetap unggul di antara semua model yang diuji.

Tabel 3. Hasil Pengujian Dengan Preprocessing

Model	TN	FP	FN	TP	Accuracy	Precision	Recall	F1-Score
Logistic Regression (preprocessing)	4,820	252	809	636	0.837	0.716	0.440	0.545
Random Forest (preprocessing)	5,034	38	418	1,027	0.930	0.964	0.711	0.818
Gradient Boosting (preprocessing)	5,015	57	441	1,004	0.924	0.946	0.695	0.801
K-Nearest Neighbors (preprocessing)	4,906	166	541	904	0.892	0.845	0.626	0.719

### 3. Hyperparameter Tuning

Model ketiga, dilakukan tuning hyperparameter pada model yang telah dipreproses untuk meningkatkan akurasi dan kinerja (Liao et al. 2022). Hasil dari tahap ini dibandingkan dengan baseline dan preprocessing untuk menentukan konfigurasi model terbaik. Hasil pengujian menunjukkan bahwa Random Forest mempertahankan keunggulannya dengan akurasi 93.0%, precision 96.4%, recall 71.1%, dan F1-score 81.8%, memperkuat posisinya sebagai model paling andal dalam keseimbangan metrik. Gradient Boosting juga tetap kompetitif dengan akurasi 92.4% dan F1-score 80.1%, menunjukkan kemampuan prediksi yang stabil. K-Nearest Neighbors (KNN) mempertahankan akurasi 89.2% dengan F1-score 71.9%, sementara Logistic Regression menunjukkan performa yang lebih rendah, dengan akurasi 83.7% dan F1-

score 54.5%. Secara keseluruhan, hyperparameter tuning berhasil mempertahankan performa unggul Random Forest, menjadikannya model terbaik untuk digunakan.

Tabel 4. Hasil Pengujian Dengan Preprocessing Dan Hyperparameter Tuning

Model	TN	FP	FN	TP	Accuracy	Precision	Recall	F1-Score
Random Forest (Tuned)	5,034	38	433	1,012	0.928	0.964	0.700	0.811
Logistic Regression (Tuned)	4,820	252	809	636	0.837	0.716	0.440	0.545
K-Nearest Neighbors (Tuned)	4,946	126	531	914	0.899	0.879	0.633	0.736
Gradient Boosting (Tuned)	5,013	59	346	1,099	0.938	0.949	0.761	0.844

### E. Evaluasi

Evaluasi dilakukan dengan membandingkan hasil dari tiga versi model (baseline, preprocessing, dan hyperparameter tuning) untuk menentukan model terbaik berdasarkan metrik akurasi, presisi, recall, dan F1-score (Imani dan Arabnia 2023). Pada tahap evaluasi model, dilakukan pengujian pada tiga versi model, yaitu baseline, preprocessing, dan hyperparameter tuning. Dari tabel 5 dapat diketahui hasil pengujian baseline menunjukkan bahwa model Random Forest memiliki performa terbaik dengan akurasi 93.0%, precision 96.4%, recall 71.1%, dan F1-score 81.8%, sementara Gradient Boosting mendekati dengan akurasi 92.4% dan F1-score 80.1%. Preprocessing data tidak memberikan perubahan signifikan dalam kinerja setiap model, dengan Random Forest tetap unggul. Namun, setelah dilakukan hyperparameter tuning, Gradient Boosting

menunjukkan peningkatan kinerja yang signifikan dengan akurasi 93.79%, precision 94.91%, recall 76.05%, dan F1-score 84.44%, melebihi performa Random Forest dalam aspek recall dan F1-score, menunjukkan model ini lebih andal dalam mendeteksi risiko gagal bayar.

Pada penelitian ini menunjukkan bahwa Gradient Boosting dengan preprocessing dan hyperparameter tuning menjadi model terbaik untuk prediksi risiko kredit pada dataset ini. Dengan peningkatan yang signifikan dalam akurasi, precision, recall, dan F1-score, terutama pada recall yang lebih tinggi, Gradient Boosting menunjukkan keunggulan dalam mendeteksi. Hasil ini mengindikasikan bahwa preprocessing dan hyperparameter tuning memberikan dampak positif terhadap sensitivitas model dengan ketidakseimbangan kelas yang tinggi.

Tabel 5. Evaluasi Hasil Pemodelan

Model	Accurac y	Pre cisio n	Rec all	F1- Sco re
Logistic Regression (Baseline)	0.83 719 5	0.71 621 6	0.44 013 8	0.54 522 1
Random Forest (Baseline)	0.93 002 9	0.96 431 9	0.71 072 7	0.81 832 7
Gradient Boosting (Baseline)	0.92 358 4	0.94 627 7	0.69 481	0.80 127 7
K-Nearest Neighbors (Baseline)	0.89 151 5	0.84 486	0.62 560 6	0.71 888 7
Logistic Regression (preprocessing)	0.83 719 5	0.71 621 6	0.44 013 8	0.54 522 1
Random Forest (preprocessing)	0.93 002 9	0.96 431 9	0.71 072 7	0.81 832 7
Gradient Boosting (preprocessing)	0.92 358 4	0.94 627 7	0.69 481	0.80 127 7
K-Nearest Neighbors (preprocessing)	0.89 151 5	0.84 486	0.62 560 6	0.71 888 7
Random Forest (Tuned)	0.92 772 7	0.96 381	0.70 034 6	0.81 112 2
Logistic Regression (Tuned)	0.83 719 5	0.71 621 6	0.44 013 8	0.54 522 1

K-Nearest Neighbors (Tuned)	0.89 918 7	0.87 884 6	0.63 252 6	0.73 561 4
Gradient Boosting (Tuned)	0.93 785 5	0.94 905	0.76 055 4	0.84 441

## KESIMPULAN

Penelitian ini menguji empat model machine learning, Logistic Regression, Random Forest, Gradient Boosting, dan K-Nearest Neighbors (KNN) untuk memprediksi risiko kredit pada dataset yang tidak seimbang. Setiap model diuji dalam tiga versi: baseline, preprocessing, dan hyperparameter tuning, menggunakan metrik evaluasi akurasi, presisi, recall, dan F1-score. Berdasarkan hasil evaluasi, Gradient Boosting dengan preprocessing dan hyperparameter tuning memberikan kinerja terbaik, mencapai akurasi 93.79%, presisi 94.91%, recall 76.05%, dan F1-score 84.44%. Proses tuning secara signifikan meningkatkan performa model, terutama dalam recall dan F1-score, yang mengindikasikan kemampuan lebih baik dalam mendeteksi kasus gagal bayar dibandingkan versi baseline atau model yang hanya melalui preprocessing.

Selain Gradient Boosting, Random Forest pada tahap baseline juga menunjukkan performa yang cukup baik, dengan akurasi 93.0% dan presisi tinggi sebesar 96.4%, meskipun tidak sebaik Gradient Boosting yang telah di-tuning. Hasil ini menegaskan bahwa preprocessing dan tuning hyperparameter berperan penting dalam peningkatan performa model. Oleh karena itu, Gradient Boosting dengan hyperparameter tuning dipilih sebagai model paling optimal untuk implementasi prediksi risiko kredit dalam penelitian ini. Model ini mampu mengelola data dengan lebih akurat dan memberikan hasil yang lebih andal, menjadikannya pilihan efektif untuk mendukung pengambilan keputusan dalam pengelolaan risiko kredit.

## REFERENSI

Bello, Oluwabusayo Adijat. 2023. "Machine Learning Algorithms for Credit Risk Assessment: An Economic and Financial Analysis." *International Journal of Management Technology* 10(1): 109–33.



- Budiman, B et al. 2023. "Machine Learning-Based Classification of House Prices: A Comparative Study." In *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, , 1–6.
- Fan, Cheng et al. 2021. "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data." *Frontiers in Energy Research* 9(March): 1–17.
- Gelar Guntara, Rangga. 2023. "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7." *Jurnal Teknologi Dan Sistem Informasi Bisnis* 5(1): 55–60.
- Imani, Mehdi, dan Hamid Reza Arabnia. 2023. "Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis." *Technologies* 11(6).
- Kulsum, Ummi, Mohamad Jajuli, dan Nina Sulistiyowati. 2022. "Analisis Sentimen Aplikasi WETV di Google Play Store Menggunakan Algoritma Support Vector Machine." *Journal of Applied Informatics and Computing* 6(2): 205–12.
- Kurniawan, Aznovri et al. 2022. "Pemilihan Metode Predictive Analytics dengan Machine Learning untuk Analisis dan Strategi Peningkatan Kualitas Kredit Perbankan." *Indonesian Journal of Applied Statistics* 5(1): 1.
- Liao, Lizhi, Heng Li, Weiyi Shang, dan Lei Ma. 2022. "An Empirical Study of the Impact of Hyperparameter Tuning and Model Optimization on the Performance Properties of Deep Neural Networks." *ACM Transactions on Software Engineering and Methodology* 31.
- Muningsih, Elly. 2022. "Kombinasi Metode K-Means Dan Decision Tree Dengan Perbandingan Kriteria Dan Split Data." *Jurnal Teknoinfo* 16(1): 113.
- Normah, Bakhtiar Rifai, Satrio Vambudi, dan Rifki Maulana. 2022. "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE." *Jurnal Teknik Komputer AMIK BSI* 8(2): 174–80. <https://ejournal.bsi.ac.id/ejournal/index.php/jtk/article/view/13041/pdf>.
- P, Krishnaraj, Rita S, dan Jitendra Jaiswal. 2024. "Comparing Machine Learning Techniques for Loan Approval Prediction."
- Raschka, Sebastian. 2018. "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning." <http://arxiv.org/abs/1811.12808>.
- Roy, Yannick et al. 2019. "Deep learning-based electroencephalography analysis: A systematic review." *Journal of Neural Engineering* 16(5).
- Shem L. Gonzales. 2023. "Transfer Learning and Tuning of Deep Pre-trained Architecture for Face Recognition." *International Journal of Advanced Research in Science, Communication and Technology*: 691–95.
- Wang, Liukai, Fu Jia, Lujie Chen, dan Qifa Xu. 2023. "Forecasting SMEs' credit risk in supply chain finance with a sampling strategy based on machine learning techniques." *Annals of Operations Research* 331(1): 1–33. <https://doi.org/10.1007/s10479-022-04518-5>.