

Penerapan Machine Learning Algoritma Random Forest untuk Prediksi Keberhasilan UMKM

Muhammad Reza Romahdoni^{*1}, Rahma Diani Sormin², Danu Budiono³

^{1,2,3} Institut Teknologi dan Bisnis Diniyah Lampung

Article History:

Received: April, 2025

Revised: Mei, 2025

Accepted: Juni, 2025

Published: Juli, 2025

Keywords: *UMKM, Machine Learning, Random Forest, Prediksi*

*Author:

m.rezarhomadoni@gmail.com¹

Abstract: *UMKM memiliki peranan penting dalam perekonomian Indonesia, khususnya dalam menciptakan lapangan kerja dan meningkatkan kesejahteraan masyarakat. Namun demikian, tidak semua UMKM mampu bertahan dan berkembang di tengah persaingan bisnis yang semakin ketat. Berbagai faktor seperti manajemen keuangan yang kurang optimal, strategi pemasaran yang lemah, hingga keterbatasan akses terhadap data dan informasi yang akurat sering kali menjadi penyebab utama kegagalan UMKM. Penelitian dilakukan melalui proses yang sistematis dan terarah guna menyelesaikan permasalahan yang dikaji. Langkah-langkah penelitian diawali dengan dataset, kemudian dilanjutkan dengan tahap preprocessing, yang bertujuan untuk membersihkan serta mempersiapkan data sebelum dilakukan analisis. Selanjutnya implementasi Random Forest, di mana algoritma diterapkan untuk membangun model prediksi, model akan melalui tahap pengujian dan evaluasi di mana performa model diukur berdasarkan akurasi. Hasil implementasi algoritma Random Forest dalam prediksi keberhasilan UMKM menunjukkan kinerja yang sangat baik dengan tingkat akurasi 86,4%, F1-score 0.862, Precision 0.861, dan Recall 0.864. Hal ini membuktikan bahwa Random Forest mampu menggeneralisasi data dengan baik dan memberikan prediksi yang andal, menjadikannya metode yang layak digunakan dalam pengembangan model prediksi keberhasilan UMKM. Beberapa aspek masih dapat ditingkatkan dalam penelitian lanjutan. Salah satu perbaikan yang dapat dilakukan adalah optimasi parameter model melalui hyperparameter tuning, seperti Grid Search atau Random Search, guna mencari kombinasi jumlah pohon keputusan dan atribut terbaik untuk meningkatkan performa prediksi. Selain itu, pengujian dengan dataset yang lebih besar.*

1. PENDAHULUAN

UMKM memainkan peran vital dalam perekonomian Indonesia. Menurut data Kementerian Koperasi dan UKM, UMKM menyumbang lebih dari 60% terhadap Produk Domestik Bruto (PDB) nasional dan menyerap sekitar 97% tenaga kerja [1]. Sesuai dengan undang-undang UMKM, semua pihak, mulai dari pemerintah pusat hingga masyarakat, bertanggung jawab untuk mengembangkan UMKM melalui berbagai dukungan seperti pelatihan, pendanaan, dan fasilitas bisnis [2].

Pendampingan Usaha Mikro, Kecil, dan Menengah (UMKM) adalah praktik membantu dan mendukung pemilik UMKM di banyak

bidang perusahaan mereka [3]. Keberhasilan usaha sendiri dapat ditandai dengan peningkatan jumlah penjualan, meningkatnya jumlah produksi, meningkatnya keuntungan atau laba serta usaha yang selalu berkembang [4].

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peranan penting dalam perekonomian Indonesia, khususnya dalam menciptakan lapangan kerja dan meningkatkan kesejahteraan masyarakat. Namun demikian, tidak semua UMKM mampu bertahan dan berkembang di tengah persaingan bisnis yang semakin ketat. Berbagai faktor seperti manajemen keuangan yang kurang optimal,

strategi pemasaran yang lemah, hingga keterbatasan akses terhadap data dan informasi yang akurat sering kali menjadi penyebab utama kegagalan UMKM

Seiring dengan perkembangan teknologi, penerapan Machine Learning mulai dilirik sebagai solusi untuk membantu UMKM dalam mengambil keputusan bisnis yang lebih tepat. Salah satu metode Machine Learning yang cukup efektif dalam proses klasifikasi dan prediksi adalah algoritma *Random Forest*. Beberapa penelitian terkait penerapan Machine Learning terkait UMKM pernah dilakukan oleh Kartika Maulida Hindrayani [5]. Hasilnya penerapan algoritma ML dapat diterapkan pada analisis sentimen, segmentasi pelanggan, prediksi harga komoditi, dan peningkatan loyalitas pelanggan. Penelitian selanjutnya dilakukan oleh Hanifah Muthia [6]. Hasil penelitian ini Machine learning terbukti mampu memberikan prediksi permintaan yang lebih akurat, mengurangi risiko stok berlebih atau kekurangan, serta memperkaya wawasan bisnis. Kemudian penelitian oleh Eugenea Chiquita Zahrani Assyarif [7]. Hasil penelitian ini diharapkan dapat membantu UMKM dalam memahami perilaku pelanggan, meningkatkan pelayanan, serta mendukung strategi pemasaran yang lebih efektif. Selanjutnya penelitian yang dilakukan oleh [8]. Implementasi algoritma Machine Learning, khususnya SARIMA sebagai Algoritma paling akurat, membantu UMKM dan toko ritel modern dalam mengantisipasi permintaan konsumen secara lebih akurat, yang berkontribusi pada optimasi manajemen persediaan, mengurangi risiko kelebihan stok atau kekurangan stok, dan meningkatkan efisiensi operasional.

Melalui penerapan algoritma *Random Forest*, diharapkan dapat dihasilkan sebuah sistem yang mampu memprediksi peluang keberhasilan UMKM berdasarkan data seperti volume penjualan, jumlah pelanggan, frekuensi transaksi, dan faktor-faktor lainnya. Prediksi ini diharapkan dapat menjadi dasar rekomendasi strategis bagi pemilik UMKM dalam merancang rencana bisnis yang lebih tepat sasaran.

Random forest merupakan metode klasifikasi yang dilakukan dengan mengembangkan metode *Decision Tree* berdasarkan pemilihan atribut acak pada setiap node untuk menentukan klasifikasi. Pada proses klasifikasinya didasarkan pada suara terbanyak dari pohon keputusan yang dikembalikan [9].

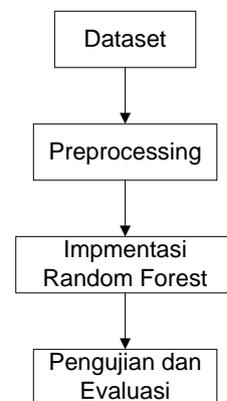
Penelitian ini dilakukan sebagai salah satu langkah dalam pemanfaatan teknologi kecerdasan buatan untuk mendukung digitalisasi sektor UMKM di Indonesia. Dengan model

prediksi yang akurat, pemerintah daerah maupun lembaga pendukung UMKM dapat memberikan intervensi dan pembinaan yang lebih efektif kepada pelaku usaha yang memiliki potensi keberhasilan tinggi maupun yang berisiko gagal.

Oleh karena itu, penelitian ini mengusung judul "Penerapan Machine Learning Algoritma *Random Forest* untuk Prediksi Keberhasilan UMKM" dengan harapan dapat memberikan kontribusi nyata dalam upaya peningkatan daya saing dan keberlanjutan bisnis UMKM.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui proses yang sistematis dan terarah guna menyelesaikan permasalahan yang dikaji. Langkah-langkah penelitian diawali dengan dataset, kemudian dilanjutkan dengan tahap preprocessing, yang bertujuan untuk membersihkan serta mempersiapkan data sebelum dilakukan analisis lebih lanjut. Selanjutnya, dilakukan implementasi *Random Forest*, di mana algoritma diterapkan untuk membangun model prediksi. Setelah itu, model akan melalui tahap pengujian dan evaluasi di mana performa model diukur berdasarkan akurasi dan efektivitasnya. Tahapan penelitian diilustrasikan dalam Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

a) Dataset

Dataset adalah sekumpulan data yang dapat digunakan sebagai bahan percobaan riset [10]. Dataset yang digunakan diperoleh dari Kaggle (<https://www.kaggle.com/>) sebagai sumberdata terbuka yang dapat diakses secara publik. Dataset ini mencakup 250 sampel dengan 13 variabel, yang mencerminkan berbagai faktor keberhasilan UMKM. Variabel-variabel tersebut meliputi umur, tingkat pendidikan pemilik UMKM,

modal awal usaha, pencatata keuangan, pemanfaatan teknologi, rencana bisnis, pemasaran, kemitraan, pengalaman bisnis, pengalaman, jenis kelamin, frekuensi dan tingkat keberhasilan.

b) *Preprocessing*

Preprocessing adalah salah satu tahapan penting dalam *text summarization*, di mana pada tahap ini teks diubah menjadi bentuk yang lebih mudah dicerna oleh komputer [10]. Analisis data dalam penelitian ini dilakukan menggunakan software *Orange*. Sebelum proses analisis, dilakukan tahap prapemrosesan data untuk memastikan bahwa dataset tersedia dalam format yang sesuai dan dapat diinterpretasikan dengan baik. Tahapan ini mencakup pembersihan data, normalisasi, serta pengolahan atribut agar hasil analisis lebih optimal dan akurat.

c) Implementasi *Random Forest*

Data yang diperoleh dari langkah sebelumnya akan diterapkan ke alat data mining, yaitu menggunakan *Orange Data Mining*. Pada tahap implementasi, langkah-langkah pengisian data ke dalam alat ini dilakukan, dan hasilnya akan menampilkan prediksi keberhasilan UMKM. *Random Forest* melakukan prediksi dengan mengombinasikan hasil dari tiap tiap pohon keputusan dengan cara *majority vote* untuk klasifikasi atau rata-rata untuk regresi [11].

d) Pengujian dan Evaluasi

Tahap pengujian dan evaluasi bertujuan untuk mengevaluasi performa *Random Forest* dalam memprediksi keberhasilan UMKM. Pada proses ini, data yang telah mengalami tahap prapemrosesan dipisahkan menjadi dua bagian utama. Hasil dari tahapan ini akan menjadi indikator utama dalam menilai sejauh mana *Random Forest* dapat dijadikan pendekatan yang efektif dalam prediksi keberhasilan UMKM.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dan pembahasan dari penelitian yang telah dilakukan, mencakup dataset, implementasi *Random Forest* serta pengujian dan evaluasi.

a) Pengumpulan Dataset

Dataset yang digunakan diperoleh dari platform Kaggle (<https://www.kaggle.com/>) dan ditampilkan dalam bentuk cuplikan layar pada Gambar 2 untuk memberikan

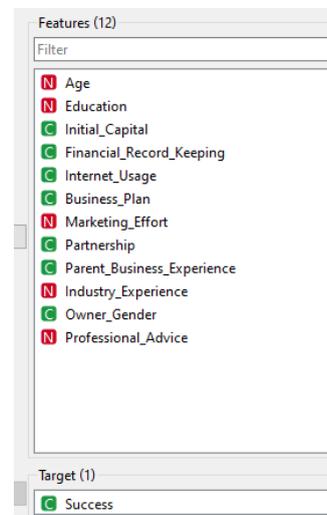
gambaran mengenai struktur data yang digunakan dalam penelitian ini.

Success	Age	Education	Initial_Capital	Initial_Record_Keeping	Internet_Usage	Business_Plan	Marketing_Effort	Partnership	Parent_Business_Experience	Industry_Experience
1	40	2.0	0	0	0	0	0	0	0	0
2	51	4.1	1	0	0	0	0	0	0	0
3	23	1.0	0	1	0	0	0	0	0	0
4	20	1.0	0	1	1	1	0	0	0	0
5	35	1.1	1	0	1	0	0	0	0	0
6	18	1.1	1	0	1	0	0	0	0	0
7	40	4.1	1	0	0	0	0	0	0	0
8	28	3.1	0	0	1	0	0	0	0	0
9	31	2.0	1	1	0	0	0	0	0	0
10	42	3.1	1	0	1	0	0	0	0	0
11	27	2.0	0	1	0	0	0	0	0	0
12	18	1.0	1	0	0	0	0	0	0	0
13	40	1.1	1	0	1	0	0	0	0	0
14	33	2.0	0	0	0	0	0	0	0	0
15	36	2.1	0	0	1	0	0	0	0	0
16	23	1.0	0	1	1	0	0	0	0	0
17	57	2.0	0	1	1	1	0	0	0	0
18	35	4.0	0	1	0	0	0	0	0	0
19	29	2.1	1	1	0	0	0	0	0	0
20	43	3.0	0	1	0	0	0	0	0	0
21	18	1.1	0	0	1	0	0	0	0	0
22	31	1.1	0	0	1	0	0	0	0	0
23	23	1.1	1	0	0	0	0	0	0	0
24	19	1.1	0	0	0	0	0	0	0	0
25	30	4.0	1	0	0	0	0	0	0	0
26	19	1.1	1	1	1	0	0	0	0	0
27	18	1.0	1	0	1	0	0	0	0	0

Gambar 2. Dataset

b) *Preprocessing*

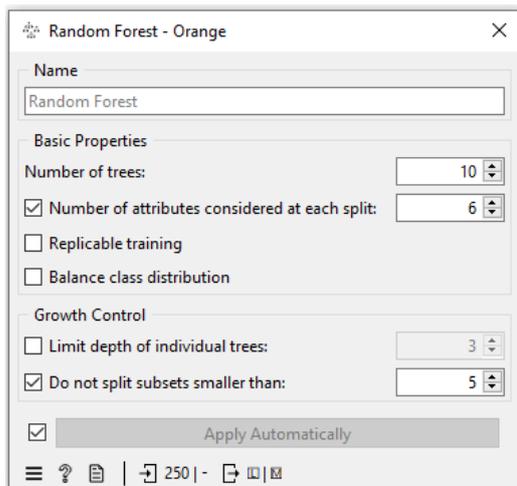
Dalam penelitian ini, dataset telah dianggap baik dan tidak mengandung nilai yang kosong (*missing value*). Tahap *preprocessing* data melibatkan pemilihan atribut yang akan digunakan sebagai atribut fitur dan atribut target. Dalam kasus ini, atribut yang diambil sebagai target adalah atribut "*Success*", sementara atribut lainnya akan dijadikan atribut fitur. Pada gambar 3 disajikan rincian dataset yang digunakan.



Gambar 3. *Features* dan *Target*

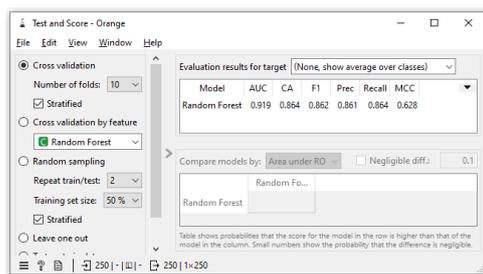
c) Implementasi *Random Forest*

Parameter pengujian yang digunakan dalam penelitian ini meliputi *Number of trees*: 10. *Number of attributes considered at each split*: 6. Detail dari parameter yang digunakan dalam algoritma *Random Forest* dapat dilihat pada Gambar 4.



Gambar 4. Parameter *Random Forest*

- d) Pengujian dan Evaluasi
 Model diuji menggunakan *Cross Validation* dengan *Number Fold* 10. Berdasarkan hasil pengujian model, diperoleh nilai AUC sebesar 0,919. Untuk detail hasil pengujian lengkap, dapat dilihat pada Gambar 5.



Gambar 5. Pengujian Model AUC

Confusion matrix/matrik konfusi yang digunakan dalam penelitian ini terdiri dari empat elemen utama, yaitu *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN). Tabel ini digunakan untuk mengevaluasi hasil prediksi model dengan membandingkan kelas aktual dan kelas yang diprediksi. Tabel 1 menyajikan hasil perhitungan confusion matrix yang digunakan dalam penelitian ini.

Tabel 1. *Confusion matrix*/matrik konfusi

Fakta	Prediksi	
	Negatif	Positif
Negatif	TN (<i>True Negative</i>)	FP (<i>False Positive</i>)
Positif	FN (<i>False Negative</i>)	TP (<i>True Positive</i>)

Terdapat beberapa metrik kinerja yang umumnya digunakan, di antaranya adalah sebagai berikut:

1. Akurasi (*Accuracy*)
 Akurasi mengindikasikan seberapa efektif model dalam mengelompokkan data dengan benar secara keseluruhan. Nilai ini diperoleh dengan menghitung rasio antara jumlah prediksi yang tepat terhadap total data yang digunakan. Persentase akurasi dalam penelitian ini dihitung menggunakan rumus berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. Presisi (*Precision*)
 Presisi (*Positive Predictive Value*) digunakan untuk menilai kinerja sistem dengan menghitung data yang diklasifikasikan dengan benar dan data yang salah diklasifikasikan. Data yang terklasifikasikan dengan benar menjadi acuan untuk memperoleh nilai presisi, dengan membaginya dengan hasil prediksi False Positive, yaitu data yang terprediksi tidak tepat. Persentase presisi dalam penelitian ini dihitung menggunakan rumus berikut.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

3. Recall
Recall adalah metode untuk mengevaluasi kinerja suatu sistem dalam mengidentifikasi kembali informasi. Ini membandingkan rasio data yang diprediksi dengan benar terhadap keseluruhan data yang sebenarnya positif

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

4. F1-Score
 F1-Score adalah nilai yang membandingkan nilai rata-rata antara presisi dan Recall. Rumus yang digunakan untuk menghitung nilai F1-Score adalah sebagai berikut

$$F1-Score = 2 \times \frac{Recall \times Precision}{Recall+Precision} \quad (4)$$

5. Nilai Area Under Curve (AUC)
 AUC (*Area Under the Curve*) mengukur luas di bawah kurva *Receiver Operating Characteristic* (ROC) dan memiliki nilai yang berkisar antara 0,5 hingga 1. Interpretasi nilai AUC terbagi dalam lima

kategori: 0,5–0,6 menunjukkan akurasi rendah, 0,6–0,7 mengindikasikan akurasi lemah, 0,7–0,8 menandakan akurasi sedang, 0,8–0,9 mencerminkan akurasi tinggi, dan 0,9–1 menggambarkan tingkat akurasi yang sangat baik

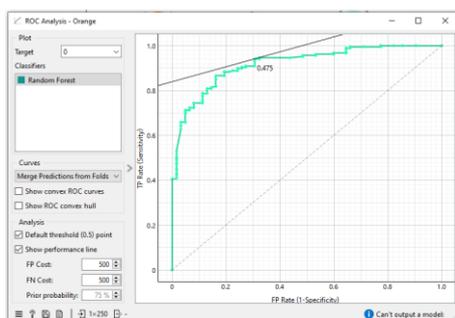
		Predicted		Σ
		0	1	
Actual	0	173	15	188
	1	19	43	62
Σ		192	58	250

Gambar 6. Matrix Konfusi

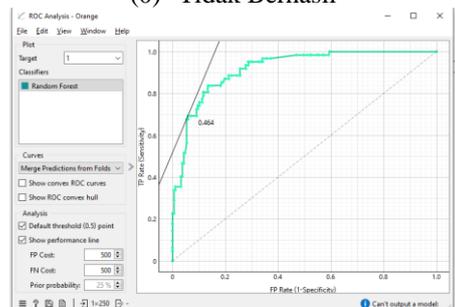
Dengan menggunakan 250 data dan menjalankan proses perulangan sebanyak 100 kali, model menghasilkan akurasi sebesar 173 untuk kategori "Tidak Berhasil" dan 43 untuk kategori "Berhasil".

6. Hasil Nilai AUC

Hasil prediksi keberhasilan UMKM divisualisasikan melalui kurva ROC, yang menunjukkan performa model dalam membedakan kategori. Gambar 7 menyajikan Kurva ROC untuk kategori "Tidak Berhasil" (0) dengan nilai 0.475 dan kategori "Berhasil" (1) dengan nilai 0.464, yang merepresentasikan tingkat keakuratan model dalam klasifikasi data



(0) Tidak Berhasil



(1) Berhasil

Gambar 7. Kurva ROC

Gambar 7.0 menampilkan kurva ROC yang menunjukkan hubungan antara *False Positive Rate* (sumbu x) dan *True Positive Rate* (sumbu y), dengan nilai AUC sebesar 0.475 sebagai indikator prediksi keberhasilan UMKM. Hasil ini menunjukkan bahwa model memiliki akurasi tinggi, karena mendekati titik 0.1. Sementara itu, Gambar 7.1 juga memperlihatkan kurva ROC dengan pola serupa, di mana nilai AUC sebesar 0.464 digunakan untuk memprediksi keberhasilan UMKM menunjukkan bahwa model tetap memiliki akurasi yang baik, dengan performa klasifikasi yang stabil.

4. KESIMPULAN

Hasil implementasi algoritma *Random Forest* dalam prediksi keberhasilan UMKM menunjukkan kinerja yang sangat baik dengan tingkat akurasi 86,4%, F1-score 0.862, *Precision* 0.861, dan *Recall* 0.864. Pengujian dilakukan dengan parameter *Number of Trees*: 10 dan *Number of Attributes Considered at Each Split*: 6 menggunakan tools *Orange Data Mining*. Hasil evaluasi berdasarkan *confusion matrix* menunjukkan bahwa model memiliki tingkat kesalahan klasifikasi yang rendah, sementara analisis kurva evaluasi yang mendekati titik 0.1 mengkonfirmasi stabilitas model dalam melakukan klasifikasi data. Hal ini membuktikan bahwa *Random Forest* mampu menggeneralisasi data dengan baik dan memberikan prediksi yang andal, menjadikannya metode yang layak digunakan dalam pengembangan model prediksi keberhasilan UMKM

Meskipun hasil penelitian ini telah menunjukkan tingkat akurasi yang tinggi, beberapa aspek masih dapat ditingkatkan dalam penelitian lanjutan. Salah satu perbaikan yang dapat dilakukan adalah optimasi parameter model melalui *hyperparameter tuning*, seperti *Grid Search* atau *Random Search*, guna mencari kombinasi jumlah pohon keputusan dan atribut terbaik untuk meningkatkan performa prediksi. Selain itu, pengujian dengan dataset yang lebih besar dan beragam diperlukan untuk memastikan bahwa model dapat beradaptasi dengan berbagai kondisi.

References

- [1] A. Farhani, A. Fitri, and R. D. Sormin, "Analisis Penerapan Ekonomi Sirkular

-
- dan Inovasi Hijau UMKM: Studi Kasus Kabupaten Pesawaran,” *MDP Student Conf.*, vol. 4, no. 2, pp. 928–935, 2025, doi: 10.35957/mdp-sc.v4i2.11219.
- [2] M. R. Romahdoni and M. A. K. Wardana, “Penerapan Business Intelligence Terhadap Strategi Pengembangan Produk Unggul Pada UMKM Ecoprint Menggunakan Algoritma Apriori,” *J. Inform.*, vol. 24, no. 2, pp. 94–107, 2024.
- [3] M. R. Romahdoni, M. Agustin, R. D. Somin, and D. Budiono, “Sistem Informasi Pendampingan UMKM Dalam Meningkatkan Kinerja Menggunakan Metode Use Case Point,” *SEAT J. Softw. Eng. Technol.*, 2024.
- [4] E. Merdekawati and N. Rosyanti, “Faktor-Faktor Yang Mempengaruhi Keberhasilan Umkm (Studi Kasus Pada Umkm Di Kota Bogor),” *JIAFE (Jurnal Ilm. Akunt. Fak. Ekon.*, vol. 5, no. 2, pp. 165–174, 2020, doi: 10.34204/jiafe.v5i2.1640.
- [5] K. Maulida Hindrayani, A. Anjani, and A. Lina Nurlaili, “Penerapan Machine Learning pada Penjualan Produk UMKM : Studi Literatur,” *Pros. Semin. Nas. Sains Data*, vol. 1, no. 01, pp. 19–23, 2021, doi: 10.33005/senada.v1i01.7.
- [6] H. Muthiah and N. K. Hamidah, “INTEGRASI MACHINE LEARNING UNTUK OPTIMALISASI PREDIKSI PERMINTAAN PRODUK PADA UMKM KULINER,” *J. PenKoMi Kaji. Pendidik. dan Ekon.*, vol. 8, no. 1, pp. 229–233, 2025.
- [7] E. C. Z. Assyari and K. D. Nuryana, “Penerapan Klasifikasi Pelanggan Berdasarkan Segmentasi Pelanggan pada UMKM Monex Toys Bekasi,” *Modem J. Inform. dan Sains Teknol.*, 2025.
- [8] A. Avinash, A. Widjaja, and O. Karnalim, “Analisis Perbandingan Algoritma Machine Learning untuk Forecasting Persediaan Produk Barang Pokok,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 361–378, 2024, doi: 10.28932/jutisi.v10i2.9357.
- [9] L. Ratnawati and D. R. Sulistyaningrum, “Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit,” *J. SAINS DAN SENI ITS*, vol. 8, no. 2, 2019.
- [10] K. U. Syaliman, “Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia,” *IT J. Res. Dev.*, vol. 5, no. 1, pp. 19–31, 2020.
- [11] M. R. Adrian, M. P. Putra, M. H. Rafialdy, N. A. Rakhmawati, and D. S. Informasi, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” *J. Inform. UPGRIS*, vol. 7, no. 1, pp. 36–40, 2021.