

**OPTIMASI PARAMETER KNN UNTUK PREDIKSI HARGA
SMARTPHONE DENGAN PENANGANAN MISSING VALUE**

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

ELISA TRI AMMAH A'FENA

NPM: 2113030069

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI

Skripsi oleh :

ELISA TRI AMMAH A'FENA

NPM : 2113030069

Judul

**OPTIMASI PARAMETER KNN UNTUK PREDIKSI HARGA
SMARTPHONE DENGAN PENGANANAN MISSING VALUE**

Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri

Tanggal: 01 Juli 2025

Pembimbing I



Sucipto, M.Kom

NIDN. 0721029101

Pembimbing II



Aidina Ristyawan, M.Kom

NIDN. 0721018801

Skripsi oleh :

ELISA TRI AMMAH A'FENA

NPM : 2113030069

Judul

**OPTIMASI PARAMETER KNN UNTUK PREDIKSI HARGA
SMARTPHONE DENGAN PENANGANAN MISSING VALUE**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 10 Juli 2025

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua Penguji : Sucipto, S.Kom, M.Kom

[.....]

2. Penguji 1 : Dwi Harini, S.Si, M.M

[.....]

3. Penguji 2 : Aidina Ristyan, S.Kom, M.Kom

[.....]



Mengetahui,
Dekan Fakultas Teknik dan Ilmu Komputer

Drs. Sunistiono, M.Si

NIDN. 0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Elisa Tri Ammah A'fena

Jenis Kelamin : Perempuan

Tempat/tgl. lahir : Kediri/ 22 Januari 2002

NPM : 2113030069

Fak/Prodi. : FTIK/ S1-Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 01 Juli 2025

Yang Menyatakan



ELISA TRI AMMAH A'FENA

NPM: 2113030069

MOTTO

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”
(QS. Al-Baqarah: 286)

”Bukan hal-hal atau peristiwa tertentu yang meresahkan kita, tetapi pertimbangan/persepsi akan hal-hal dan peristiwa tersebut.”
(Epictetus-Filosofi Teras)

“Berhasil bukan hanya mendapatkan apa yang direncanakan, tapi berhasil adalah mampu bangkit dari hal yang tidak diharapkan”
(Anies Baswedan)

“Hidup memang tak selalu mudah, tapi selalu ada kekuatan dalam setiap luka yang kita sembuhkan sendiri.”
(Najwa Shihab)

“Tidak semua hal harus dimiliki. Tidak semua yang pergi harus dikejar. Sebab hidup, pada akhirnya, adalah tentang menerima.”
(Fiersa Besari-Konspirasi Alam Semesta)

“Setinggi apapun ilmu yang kita miliki, semoga membuat hati kita selalu membumi”
(Karis)

ABSTRAK

Elisa Tri Ammah A'fena: *K-Nearest Neighbors* Untuk Menangani *Missing Value* Pada Prediksi Harga *Smartphone*, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2025.

Kata Kunci: Prediksi, Harga, *Smartphone*, *K-Nearest Neighbors*, *Missing Value*, *Regressi*.

Perkembangan teknologi digital mendorong meningkatnya kebutuhan masyarakat terhadap perangkat *smartphone* sebagai alat komunikasi dan pusat aktivitas *modern*. Seiring meningkatnya permintaan, penentuan harga *smartphone* yang tepat menjadi tantangan bagi pelaku usaha karena dipengaruhi berbagai faktor seperti spesifikasi teknis, tren pasar, serta kondisi data yang seringkali tidak lengkap. Ketidaklengkapan data atau *missing value* dapat menurunkan akurasi prediksi harga, sehingga dibutuhkan metode penanganan yang tepat agar hasil prediksi tetap *relevan* dan mendukung pengambilan keputusan bisnis yang kompetitif.

Penelitian ini bertujuan untuk mengetahui pengaruh perbedaan metode *imputasi* data, yaitu *median* dan *KNNImputer*, terhadap *performa* algoritma *K-Nearest Neighbors* dalam memprediksi harga *smartphone*. Penelitian dilakukan dengan pendekatan CRISP-DM yang meliputi tahapan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, hingga *Deployment*. *Dataset* yang digunakan diperoleh dari platform *Kaggle*, dengan proses pra-pemrosesan meliputi *imputasi* data hilang, *feature selection*, *scaling*, dan validasi model menggunakan *cross-validation 5-fold*.

Hasil penelitian menunjukkan bahwa baik metode *imputasi median* maupun *KNNImputer* menghasilkan *performa* model KNN yang identik dengan nilai R^2 pada data uji sebesar 0,9437, MAE sebesar 57,48, serta MSE sekitar 7.543,66. *Parameter* optimal diperoleh pada konfigurasi $n_neighbors = 7$, $p = 2$, dan bobot *distance*. Meski demikian, nilai R^2 yang tinggi pada data pelatihan (1,0000) dibandingkan dengan data uji mengindikasikan potensi *overfitting* yang perlu diantisipasi pada pengembangan model lanjutan.

Secara umum, penelitian ini menegaskan bahwa strategi *imputasi missing value* dapat memengaruhi kelengkapan data, tetapi tidak selalu menghasilkan perbedaan signifikan pada akurasi model prediksi jika *parameter* KNN dioptimasi dengan tepat. Temuan ini diharapkan dapat menjadi acuan bagi pelaku bisnis dalam merancang strategi penetapan harga *smartphone* yang lebih tepat sasaran serta menjadi referensi bagi peneliti selanjutnya dalam pengembangan model *machine learning* berbasis kualitas data yang optimal.

KATA PENGANTAR

Puji Syukur Kami panjatkan Kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul “Optimasi Parameter KNN Untuk Prediksi Harga Smartphone Dengan Penanganan Missing Value” ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri. Pada kesempatan ini diucapkan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak Dr. Zainal Afandi, M.Pd. selaku Rektor UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
2. Bapak Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
3. Bapak Sucipto, S.Kom, M.Kom. selaku Ka Prodi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri serta selaku Dosen Pembimbing yang dengan sabar membimbing dan memotivasi dalam menyelesaikan tugas akhir ini.
4. Bapak Aidina Ristyawan, S.Kom, M.Kom. selaku Pembimbing skripsi yang dengan sabar membimbing dan memotivasi dalam menyelesaikan skripsi ini.
5. Terima kasih sebesar-besarnya kepada kedua orang tua atas kasih sayang perhatian, dan do’a yang tidak pernah putus. Segala bentuk dukungan, pengorbanan, dan usaha yang menjadi sumber kekuatan utama dalam setiap langkah penulis, hingga akhirnya penulis dapat menyelesaikan tugas akhir ini.
6. Terima kasih kepada kedua kakak dan tante yang telah mendukung dalam do’a, dana dan motivasi dalam penyelesaian tugas akhir ini.
7. Terima kasih kepada Arinda Sekar Bilbina Suryadi, sahabat penulis, atas segala bentuk motivasi, dukungan, dan ketulusan yang diberikan. Terima kasih telah menjadi teman yang selalu hadir, menemani dalam suka maupun duka, serta membantu penulis melewati masa-masa sulit.

8. Terima kasih kepada Ovelina Devi Kurnia, teman baik penulis, atas kebersamaan, dukungan, dan bantuan yang telah diberikan selama masa perkuliahan. Terima kasih telah menjadi sosok yang hadir tidak hanya dalam proses akademik, tetapi juga dalam situasi pribadi yang sulit, serta bersedia menemani dan memberikan semangat di saat-saat terberat.
9. Terima kasih kepada Atma Agilia Triwardani, teman baik penulis yang selalu siap membantu dalam berbagai hal selama proses penyusunan tugas akhir ini. Terima kasih juga atas masukan dan semangat yang diberikan, terutama di masa-masa sulit. Bisa berteman dan berjalan bersama dengan sosok yang sefrekuensi seperti Atma adalah salah satu hal yang paling penulis syukuri dalam perjalanan ini.
10. Terima kasih kepada teman-teman kos tercinta: Putri, Nafis, dan Amel, atas perhatian, dukungan, dan kebersamaan yang hangat selama masa penyusunan tugas akhir ini. Kehadiran kalian benar-benar membantu dan membuat hari-hari terasa lebih ringan dan menyenangkan.
11. Terima kasih kepada diri sendiri yang telah bertahan dan berjuang keras hingga akhirnya tugas akhir ini dapat diselesaikan. Banyak hal yang telah dilewati dalam proses ini, tangis yang tak terhitung, tawa yang menguatkan. Terima kasih telah terus melangkah, tetap bertanggung jawab atas mimpi dan harapan, meskipun jalannya sering kali terasa berat.
12. Ucapan terima kasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan Skripsi ini.

Disadari bahwa skripsi ini masih banyak kekurangan, maka diharapkan kritik dan saran, dari berbagai pihak sangat diharapkan. Akhirnya, disertai harapan semoga skripsi ini ada manfaatnya bagi kita semua, khususnya bagi dunia pendidikan.

Kediri, 01 Juli 2025

ELISA TRI AMMAH A'FENA

NPM: 2113030069

DAFTAR ISI

Halaman Sampul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan.....	iv
Halaman Motto.....	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Batasan Masalah.....	3
C. Rumusan Masalah	4
D. Tujuan Penelitian	4
E. Manfaat Penelitian	4
BAB II KAJIAN PUSTAKA	5
A. Kajian Teori	5
1. Data mining.....	5
2. Prediksi.....	5
3. Harga.....	5
4. Machine Learning	6
5. Algoritma K-Nearest Neighbor.....	6

6. Imputasi Data	7
B. Kajian Hasil Penelitian Terdahulu	7
C. Kerangka	9
D. Hipotesis.....	10
BAB III METODE PENELITIAN.....	11
A. Business Understanding	11
B. Data Understanding.....	12
C. Data Preparation.....	14
D. Modeling	16
E. Evaluation	17
F. Deployment.....	19
BAB IV HASIL DAN PEMBAHASAN	20
A. Hasil	20
B. Pembahasan.....	22
1. Business Understanding	24
2. Data Understanding.....	24
3. Data Preparation	56
4. Modelling	63
5. Evaluasi	67
6. Deployment.....	70
BAB V PENUTUP.....	72
A. Kesimpulan	72
B. Saran.....	72
DAFTAR PUSTAKA	73
LAMPIRAN.....	80

DAFTAR GAMBAR

Gambar 3.1 Tahapan Metode CRISP-DM	11
Gambar 4.1 Visualisasi Hasil Evaluasi	21
Gambar 4.2 Visualisasi Fitur phone_model	25
Gambar 4.3 Hasil Fitur Store	27
Gambar 4.4 Visualisasi Fitur Price	28
Gambar 4.5 Visualisasi Fitur Price_USD	29
Gambar 4.6 Visualisasi price range	34
Gambar 4.7 Hasil Fitur Storage	35
Gambar 4.8 Visualisasi Fitur RAM	36
Gambar 4.9 Visualisasi Fitur Battery	37
Gambar 4.10 Visualisasi Fitur Chipset	38
Gambar 4.11 Visualisasi Fitur CPU	39
Gambar 4.12 Visualisasi Fitur GPU	40
Gambar 4.13 Visualisasi Fitur Foldable	41
Gambar 4.14 Visualisasi Fitur Display_Type	42
Gambar 4.15 Visualisasi Fitur Display Size	43
Gambar 4.16 Visualisasi Fitur Display Resolution	44
Gambar 4.17 visualisasi sasi Fitur PPI_Density	45
Gambar 4.18 Visualisasi Fitur Dimensions	46
Gambar 4.19 Visualisasi Fitur Weight	47
Gambar 4.20 Visualisasi Fitur OS	48
Gambar 4.21 Visualisasi Fitur NFC	49
Gambar 4.22 Visualisasi Fitur USB	50
Gambar 4.23 Visualisasi Feature sensor	51
Gambar 4.24 Visualisasi Fitur Colors	52
Gambar 4.25 Visualisasi Fitur Video	53
Gambar 4.26 Visualisasi Fitur Launch	55
Gambar 4. 27 Visualisasi Fitur Year	56

DAFTAR TABEL

Tabel 2.1 Rumus KNN.....	7
Tabel 3.2 Rumus MAE	17
Tabel 3.3 Rumus MSE	18
Tabel 3.4 Rumus R^2	18
Tabel 4.1 Hasil Evaluasi	20
Tabel 4.2 Hasil Fitur Phone Model	26
Tabel 4.3 Hasil Fitur Quantil_10	30
Tabel 4.4 Hasil Fitur Quantil_50	32
Tabel 4.5 Hasil Fitur Fitur Quantil_90.....	32
Tabel 4.6 Cek missing value	57
Tabel 4.7 Tabel fitur phone_brand, store, price, currency, storage, ram	60
Tabel 4.8 Tabel fitur Launch, Weight, Display_Type, Display_Size, NFC, USB	60
Tabel 4.9 Tabel fitur Display_Size, NFC, USB.....	61
Tabel 4.10 Tabel fitur Battery, year, foldable, PPI_Density, quantile_10.....	61
Tabel 4.11 Tabel fitur quantile_50, quantile_90, price_range	61
Tabel 4.12 Tabel fitur baris label (y)	61
Tabel 4.13 Tabel baseline	64
Tabel 4.14 Data yang akan diuji	65
Tabel 4.15 Jumlah tetangga.....	66
Tabel 4.16 Variasi Parameter	66

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Skripsi	80
Lampiran 2. Surat Bebas Similarity	81
Lampiran 3. Bukti Halaman Awal cek Similarity.....	82
Lampiran 6. Bukti LOA Publikasi Jurnal	84
Lampiran 7. Code Imputasi Median	93
Lampiran 8. Code Imputasi KNNImputer	100

BAB I

PENDAHULUAN

A. Latar Belakang

Di era digital dan *modern* saat ini, sektor perdagangan yang banyak mengalami perkembangan. Perkembangan ini juga memberikan dampak besar terhadap industri teknologi. Berbagai produk teknologi pun mengalami peningkatan permintaan dan menjadi kebutuhan pokok dalam kehidupan digital. Salah satunya adalah *smartphone*, yang kini berperan sebagai pusat aktivitas komunikasi dan teknologi modern.

Teknologi komunikasi saat ini memegang peran yang sangat penting, karena meningkatnya kebutuhan akan pertukaran informasi yang cepat dan akurat. Kemajuan dalam teknologi komunikasi membuat manusia dapat saling terhubung tanpa hambatan jarak, ruang, atau waktu (Setiawan & Nugroho, 2021). *Smartphone* sendiri merupakan perangkat yang tidak hanya digunakan untuk berkomunikasi, tetapi juga dilengkapi dengan fungsi *Personal Digital Assistant* (PDA) dan memiliki kemampuan layaknya sebuah komputer (Green Ferry Mandias, 2017). Seiring meningkatnya ketergantungan masyarakat terhadap *smartphone*, prediksi harga *smartphone* menjadi penting dan dipengaruhi oleh berbagai faktor, seperti spesifikasi teknis, tren pasar, serta kondisi sosial dan ekonomi global.

Preferensi konsumen juga berperan penting dalam menentukan arah harga. Ketika tren pasar beralih ke teknologi baru seperti jaringan 5G, layar lipat, atau AI pada *smartphone*, maka perangkat dengan fitur-fitur tersebut akan mengalami kenaikan harga karena meningkatnya permintaan. Namun, tidak semua konsumen mengejar fitur terbaru, sebagian besar pengguna lebih mengutamakan keseimbangan antara harga dan fungsi. Hal ini menciptakan segmen pasar yang berbeda, mulai dari pengguna yang mengejar teknologi *premium* hingga mereka yang lebih memilih perangkat ekonomis untuk kebutuhan dasar. Kondisi ini menyebabkan produsen harus menyesuaikan strategi harga mereka agar dapat menjangkau seluruh lapisan pasar. Hal ini

menjadi salah satu tantangan yang sering dihadapi oleh pelaku usaha bagaimana cara menentukan harga jual yang pas.

Salah satu kendala utama dalam menentukan harga yang akurat adalah keterbatasan informasi yang dimiliki tentang suatu produk (Wigati et al., 2024). Terkadang, informasi yang dimiliki tentang suatu produk tidak selalu lengkap, entah karena data teknis tidak tersedia sepenuhnya. Ketidaklengkapan informasi ini dapat menyebabkan penentuan harga menjadi kurang tepat dan berisiko merugikan bisnis (Ashary et al., 2023). Untuk mengatasi tantangan ini, dibutuhkan pendekatan yang mampu mengolah informasi yang tersedia secara optimal dan tetap memberikan prediksi yang akurat meskipun ada kekurangan data. Salah satu pendekatan yang kini banyak digunakan adalah pemanfaatan teknologi *Machine Learning*. *Machine Learning* sendiri adalah cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dan mengambil keputusan berdasarkan data, tanpa harus diprogram secara *eksplisit* untuk setiap tugas (Kurniati & Sulistiawati, 2023). Dalam konteks prediksi harga *smartphone*, *Machine Learning* dapat mengolah berbagai informasi yang *kompleks* dan beragam. Di sinilah peran teknologi *Machine Learning* menjadi sangat penting sebagai alat bantu dalam *prediksi* harga *smartphone* di era digital (Ryan Pratama et al., 2025).

Dalam penelitian ini, menggunakan dua metode untuk mengisi data yang hilang, yaitu *median* (nilai tengah) dan *KNNimputer*. Pendekatan ini memungkinkan model prediksi untuk bekerja dengan data yang lebih lengkap, sehingga hasil prediksi harga menjadi lebih realistis dan dapat diandalkan. Dengan menggabungkan kekuatan teknologi *Machine Learning* dan teknik *imputasi* data, pelaku bisnis dapat memperoleh gambaran harga *smartphone* yang lebih akurat terhadap perubahan pasar. Algoritma *K-Nearest Neighbor* (KNN) juga digunakan dalam penelitian ini. Algoritma ini dikenal efektif dalam menangani data *numerik* dan *kategorikal* (Afdal et al., 2024).

Penanganan *missing value* adalah salah satu masalah dalam proses *preprocessing*. Karena mudah dan cepat, metode *imputasi median* digunakan pada tahap awal. Selain itu, *median* dipilih karena tahan terhadap *outlier*

(Zahrotul Kamalia et al., 2025). Selain menggunakan *median*, penelitian ini juga menerapkan metode *KNNImputer* sebagai alternatif untuk menangani *missing*. *KNNImputer* dipilih karena dapat bekerja dengan cara mengisi *missing* berdasarkan kemiripan dengan data lain yang paling dekat. Pendekatan ini membuat hasil *imputasi* menjadi lebih akurat. Penelitian sebelumnya juga, menunjukkan bahwa penggunaan *KNNImputer* dapat meningkatkan akurasi model (Ristyawan et al., 2025). Dibandingkan dengan metode sederhana seperti *median*, *KNNImputer* memberikan nilai *imputasi* yang lebih dinamis karena menggunakan prinsip tetangga terdekat. Oleh karena itu, tujuan penelitian ini adalah membandingkan metode *imputasi median* dan *KNNImputer* untuk mengetahui dampak terhadap kinerja model dalam memprediksi harga *smartphone*. Proses *preprocessing* sangat penting karena hasil akhir pemodelan dapat sangat dipengaruhi oleh nilai-nilai yang hilang.

Hasil penelitian ini diharapkan dapat membantu dalam memilih *strategi preprocessing* yang tepat, terutama untuk menangani *missing value* serta membantu pelaku bisnis menentukan harga yang tepat untuk menentukan harga *smartphone*. Selain itu, temuan ini juga diharapkan dapat berguna bagi peneliti selanjutnya yang ingin mengoptimalkan kinerja model *machine learning* dengan mempertimbangkan elemen kualitas data sebelum pemodelan. Temuan ini juga membuka peluang penerapan metode serupa pada dataset dengan karakteristik berbeda untuk melihat konsistensi performa. Dengan demikian, rekomendasi metode imputasi dapat lebih bermanfaat secara praktis.

B. Batasan Masalah

Batasan masalah dalam memprediksi harga *smartphone* adalah:

1. Data yang digunakan dalam proses penelitian ini diambil dari *Kaggle*
<https://www.kaggle.com/datasets/willianoliveiragibin/phone-information-2024>
2. Algoritma yang penulis gunakan adalah *K-Nearest Neighbor*.
3. *Software* yang digunakan penulis adalah *jupyter notebook* dengan bahasa pemrograman *python*.

4. Penelitian ini hanya difokuskan pada metode *imputasi* untuk menangani data yang hilang

C. Rumusan Masalah

Berdasarkan penjelasan diatas, rumusan masalah yang akan dipecahkan adalah sebagai berikut:

1. Bagaimana pengaruh metode *imputasi* data yang berbeda (*median* dan KNN *imputer*) terhadap akurasi prediksi harga *smartphone* menggunakan algoritma *K-Nearest Neighbor*?
2. *Imputasi* mana yang lebih efektif dalam meningkatkan *performa* model KNN dalam prediksi harga *smartphone*?

D. Tujuan Penelitian

Tujuan dari penelitian ini antara lain:

1. Mengetahui pengaruh perbedaan metode *imputasi* data terhadap akurasi prediksi harga *smartphone*
2. Mengetahui *imputasi* mana yang lebih efektif dalam meningkatkan *performa* terhadap model.

E. Manfaat Penelitian

Manfaat dari penelitian ini dapat memberikan pemahaman metode *imputasi* mana yang paling efektif untuk menangani *missing value*. Pemilihan *imputasi* yang tepat dapat meningkatkan akurasi model, serta bermanfaat untuk pengambilan keputusan, baik bagi penjual maupun konsumen, serta memberikan keuntungan dalam mengatasi tantangan persaingan persaingan bisnis yang semakin ketat. Selain itu, penelitian ini juga berpotensi untuk mengembangkan *strategi* harga yang lebih efektif pada Toko *smartphone*, sehingga dapat meningkatkan daya saing di pasar. Hasil ini diharapkan dapat menjadi acuan bagi penelitian selanjutnya agar dapat mengembangkan metode yang lebih inovatif.

DAFTAR PUSTAKA

- Afdal, M., Novita, R., Studi Sistem Informasi, P., & Sains dan Teknologi, F. (2024). Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 10–19. <https://doi.org/10.57152/MALCOM.V4I1.983>
- Aini, E. D. N., Khasanah, R. A., **Ristyawan, A., & Diniati, E.** (2024). Penggunaan Data Mining untuk Prediksi tingkat Obesitas di Meksiko Menggunakan Metode Random Forest. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8(3), 1256–1265. <https://doi.org/10.29407/EP0PZY43>
- Andia, R., Kaslani, K., Permana, S. E., & Handayani, T. (2024). PERAMALAN HASIL PANEN PADI KABUPATEN CIREBON MENGGUNAKAN ALGORITMA REGRESI LINEAR BERGANDA. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 738–747. <https://doi.org/10.36040/JATI.V8I1.8446>
- Ashary, L., Ali, Abd. H., & Muslim, A. A. (2023). Analisis Harga Pokok Produksi Untuk Penentuan Harga Jual Es Balok di Pabrik Es Ma'un Baarid Sukorejo Situbondo. *Mazinda : Jurnal Akuntansi, Keuangan, Dan Bisnis*, 1(1), 72–82. <https://doi.org/10.35316/MAZINDA.V1I1.2868>
- Asmara, R. A., Prasetyo, A., Stevani, S., Hapsari, R. I., Studi, P., Informatika, T., Informasi, J. T., & Malang, P. N. (2021). Prediksi Banjir Lahar Dingin pada Lereng Merapi menggunakan Data Curah Hujan dari Satelit. *Jurnal Informatika Polinema*, 7(2), 35–42. <https://doi.org/10.33795/JIP.V7I2.494>
- Chazar, C., & Widhiaputra, B. E. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 12(1), 67–80. <https://doi.org/10.37424/INFORMASI.V12I1.48>
- Deepak, P. L., & Jagadeesh, P. (2024). Performance Analysis of Novel Linear Regression Algorithm with Improved Accuracy Compared over K-Nearest Neighbor in Predicting Wind Power Generation. *E3S Web of Conferences*, 491, 02043. <https://doi.org/10.1051/E3SCONF/202449102043>
- Desti, Sari, P., Shofia Hilabi, S., & Hananto, A. (2023). Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama. *SMARTICS Journal*, 9(1), 14–19. <https://doi.org/10.21067/SMARTICS.V9I1.8088>
- Fadillah, I. J., Dwi Puspita, C., & Statistik, B. P. (2020). PEMANFAATAN METODE WEIGHTED K-NEAREST NEIGHBOR IMPUTATION

(WEIGHTED KNNI) UNTUK MENGATASI MISSING DATA. *Seminar Nasional Official Statistics*, 2020(1), 511–518. <https://doi.org/10.34123/SEMNASOFFSTAT.V2020I1.409>

Febriansyah, F., & Muntari, S. (2023). Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 8(1), 66–77. <https://doi.org/10.14421/jiska.2023.8.1.66-77>

Green Ferry Mandias. (2017). Analisis Pengaruh Pemanfaatan Smartphone Terhadap Prestasi Akademik Mahasiswa Fakultas Ilmu Komputer Universitas Klabat. *CogITO Smart Journal*, 3(1), 83–90. <https://doi.org/10.31154/COGITO.V3I1.47.83-90>

Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning. *JBASE - Journal of Business and Audit Information Systems*, 4(2), 16–22. <https://doi.org/10.30813/JBASE.V4I2.3000>

Hamidah, K., Voutama, A., Artikel, S., Faktor, K. :, Kdd, K., & Linear, R. (2023). Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data World Happiness Report dengan Metode Regresi Linier. *Explore IT: Jurnal Keilmuan Dan Aplikasi Teknik Informatika*, 15(1), 1–7. <https://doi.org/10.35891/EXPLORIT.V15I1.3874>

Haryati, S., Sudarsono, A., & Suryana, E. (2015). IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: UNIVERSITAS DEHASSEN BENGKULU). *JURNAL MEDIA INFOTAMA*, 11(2). <https://doi.org/10.37676/JMI.V11I2.260>

Indarwati, T., Irawati, T., & Rimawati, E. (2019). PENGGUNAAN METODE LINEAR REGRESSION UNTUK PREDIKSI PENJUALAN SMARTPHONE. *Jurnal Teknologi Informasi Dan Komunikasi (TIKoSIN)*, 6(2). <https://doi.org/10.30646/TIKOMSIN.V6I2.369>

Insany, G. P., Yustiana, I., & Rahmawati, S. (2023). Penerapan KNN dan ANN pada klasifikasi status gizi balita berdasarkan indeks antropometri. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(2), 385–393. <https://doi.org/10.37859/COSCITECH.V4I2.5079>

Kristanto, D., Ipmawan, H., Kuncoro, A. W., Naryoto, P., Pakpahan, M., Hendrawan, K., Budi, U., & Jakarta, L. (2023). Pengaruh Kualitas Produk, Promosi Penjualan dan Ekuitas Merek Terhadap Keputusan Pembelian (Studi Pada Konsumen Smartphone Samsung di Jakarta). *Jurnal Bintang Manajemen*, 1(2), 84–97. <https://doi.org/10.55606/JUBIMA.V1I2.1382>

Kurnia, O. D., A'Fena, E. T., Ningrum, D. Y. A., **Daniati, E., & Ristyawan, A.** (2024). Analisis Perbandingan Algoritma Naïve Bayes dengan K-Nearest Neighbor (KNN) Pada Dataset Mobile Price Classification. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8(2), 1174–1183. <https://doi.org/10.29407/INOTEK.V8I2.5053>

Kurniati, I., & Sulistiawati, S. (2023). PEMANFAATAN MACHINE LEARNING UNTUK PRICE OPTIMATION DENGAN MENGGUNAKAN MODEL ARTIFICIAL NEURAL NETWORK. *JEIS: Jurnal Elektro Dan Informatika Swadharma*, 3(2), 83–93. <https://doi.org/10.56486/JEIS.VOL3NO2.363>

Kurniawan, R., Mulyani, Y., Budi Wintoro, P., Komarudin, M., Informatika, T., Lampung, U., Jl Ir Sumantri Brojonegoro No, J., & Lampung, B. (2023). IMPLEMENTASI ARSITEKTUR XCEPTION PADA MODEL MACHINE LEARNING KLASIFIKASI SAMPAH ANORGANIK. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(2), 2830–7062. <https://doi.org/10.23960/JITET.V11I2.3034>

Lee, H., & Yun, S. (2024). Strategies for Imputing Missing Values and Removing Outliers in the Dataset for Machine Learning-Based Construction Cost Prediction. *Buildings*, 14(4). <https://doi.org/10.3390/BUILDINGS14040933>

Liyadi, K. R., Pratiwi, H., Aditya, P., & Sa'ad, M. I. (2022). Penerapan Metode Single Moving Average Dalam Peramalan Persediaan Bahan Pangan. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 4(1), 72–80. <https://doi.org/10.30645/BRAHMANA.V4I1.136>

Maharani, M., & Fathoni, F. (2024). Analisis Sentimen Pengguna Terhadap Faktor Penggunaan PayPal Menggunakan Metode Decision Tree. *Jurnal Ilmiah Teknologi Informasi Asia*, 18(1), 71–83. <https://doi.org/10.32815/JITIKA.V18I1.1002>

Marlius, D., & Jovanka, N. (2023). PENGARUH HARGA DAN PROMOSI TERHADAP KEPUTUSAN PEMBELIAN KONSUMEN Y.O.U PADA HASANAH MART AIR HAJI. *JURNAL ECONOMINA*, 2(2), 476–490. <https://doi.org/10.55681/ECONOMINA.V2I2.326>

Mughnyanti, M., Hafiz, S., & Ginting, N. (2023). Data Mining Manhattan Distance dan Euclidean Distance Pada Algoritma X-Means Dalam Klasifikasi Minat dan Bakat Siswa. *REMIK: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 7(1), 835–842. <https://doi.org/10.33395/REMIK.V7I1.12162>

Muzaki, M. R., Nurfajriana, I. M., Ilahi, F. P. A., **Ristyawan, A., & Daniati, E.** (2024). Implementasi Data Mining dengan Algoritma K-Nearest Neighbors

untuk Memprediksi Risiko Diabetes Menggunakan Chatbot Telegram. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8(1), 310–318. <https://doi.org/10.29407/INOTEK.V8I1.4946>

Nugroho, A., & Amrullah, A. (2023). EVALUASI KINERJA ALGORITMA K-NN MENGGUNAKAN K-FOLD CROSS VALIDATION PADA DATA DEBITUR KSP GALIH MANUNGGA. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 5(2), 294–300. <https://doi.org/10.51401/JINTEKS.V5I2.2506>

Nur Aziza, ah, Yuli Astuti, R., Akbar Maulana, B., & Hidayati, N. (2024). Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Ketahanan Pangan di Provinsi Jawa Tengah. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 404–412. <https://doi.org/10.57152/MALCOM.V4I2.1201>

Nurfauzi, Y., Taime, H., Hanafiah, H., Yusuf, M., & Asir, M. (2023). Literature Review: Analisis Faktor yang Mempengaruhi Keputusan Pembelian, Kualitas Produk dan Harga Kompetitif. *Management Studies and Entrepreneurship Journal (MSEJ)*, 4(1), 183–188. <https://doi.org/10.37385/MSEJ.V4I1.1246>

Pangestu, M. S., & Fitriani, M. A. (2022). Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means. *Sainteks*, 19(2), 141–155. <https://doi.org/10.30595/SAINTEKS.V19I2.14495>

Permana, A. P., Ainiyah, K., & Holle, K. F. H. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(3), 178–188. <https://doi.org/10.14421/JISKA.2021.6.3.178-188>

Petrazzini, B. O., Naya, H., Lopez-Bello, F., Vazquez, G., & Spangenberg, L. (2021). Evaluation of different approaches for missing data imputation on features associated to genomic data. *BioData Mining*, 14(1), 1–13. <https://doi.org/10.1186/S13040-021-00274-7/FIGURES/3>

Pratiwi, G., & Lubis, T. (2021). PENGARUH KUALITAS PRODUK DAN HARGA TERHADAP KEPUASAN PELANGGAN UD ADLI DI DESA SUKAJADI KECAMATAN PERBAUNGAN. *Jurnal Bisnis Mahasiswa*, 1(2), 121–134. <https://doi.org/10.60036/JBM.V1I2.11>

Putri, A. E., MZ, Y., & Bororing, J. E. (2025). IMPLEMENTASI K-MEANS CLUSTERING DAN MODEL CRISP-DM UNTUK PENGELOMPOKAN DAERAH RAWAN TINDAK PIDANA NARKOBA DI DIY. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 7(2), 606–615. <https://doi.org/10.51401/JINTEKS.V7I2.5690>

Ravega, Putri, W., **Ristyawan, A., & Muzaki, M. N.** (2022). Comparison Performance of K-NN and NBC Algorithm for Classification of Heart Disease. *Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem Dan Komputer*, 2(2), 143–154. <https://doi.org/10.32503/JTECS.V2I2.2708>

Rekayasa, K. K., Awaliyah, C. A., Prasetyadi, A., & Junaidi, A. (2022). Sistem Rekomendasi Desain Website Berdasarkan Tingkat Kemiripan Menggunakan Euclidean Distance. *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, 2(2), 75–81. <https://doi.org/10.20895/DINDA.V2I2.543>

Rifai, A., Hidayat, M., & Nulngafan, N. (2025). PREDIKSI HARGA KENTANG DI WONOSOBO DENGAN MENGGUNAAN METODE LONG SHORT TERM MEMORY (LSTM). *Biner: Jurnal Ilmiah Informatika Dan Komputer*, 4(1), 9–18. <https://doi.org/10.32699/BINER.V4I1.7794>

Rino Pratama, R., & Korespondensi, A. (2020). Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(2), 302–311. <https://doi.org/10.30812/matrik.v19i2.688>

Ristyawan, A., Nugroho, A., & Amarya, T. K. (2025). Optimasi Preprocessing Model Random Forest untuk Prediksi Stroke. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 12(1), 29–44. <https://doi.org/10.35957/JATISI.V12I1.9587>

Ryan Pratama, A., Wabula, F., Ilmandry, H., Laura Isabela, M., Raharjo, M., Sianipar, R., Kampus, A., Raya Puspitak, J., Pamulang, K., & Tangerang Selatan, K. (2025). Literature Review The Impact of Machine Learning in Modern Industries. *Nian Tana Sikka: Jurnal Ilmiah Mahasiswa*, 3(1), 177–182. <https://doi.org/10.59603/NIANTANASIKKA.V3I1.680>

Santiastry, S., Asriyanik, A., & Apriandari, W. (2024). PENERAPAN ALGORITMA NAIVE BAYES DAN METODE CRISP-DM DALAM PREDIKSI HASIL TES KEMAMPUAN BAHASA INGGRIS MAHASISWA. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 10432–10439. <https://doi.org/10.36040/JATI.V8I5.11069>

Setiawan, R., & Nugroho, W. (2021). Rancang Bangun Aplikasi Pemesanan Makanan Berbasis Android. *Journal of Information Systems and Informatics*, 3(2), 329–340. <https://doi.org/10.33557/JOURNALISI.V3I2.132>

Sijabat, P. I., Yuhandri, Y., Nurcahyo, G. W., & Sindar, A. (2020). Algoritma Backpropagation Prediksi Harga Komoditi terhadap Karakteristik Konsumen Produk Kopi Lokal Nasional. *Digital Zone: Jurnal Teknologi*

Informasi Dan Komunikasi, 11(1), 96–107.
<https://doi.org/10.31849/DIGITALZONE.V11I1.3880>

Siregar, A. M., Faisal, S., Cahyana, Y., & Priyatna, B. (2020). Perbandingan Algoritme Klasifikasi Untuk Prediksi Cuaca. *Jurnal Accounting Information System (AIMS)*, 3(1), 15–24. <https://doi.org/10.32627/AIMS.V3I1.280>

Sucipto, S., Dwi Prasetya, D., & Widiyaningtyas, T. (2024). Educational Data Mining: Multiple Choice Question Classification in Vocational School. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 23(2), 379–388. <https://doi.org/10.30812/matrik.v23i2.3499>

Sumarsid Sumarsid, & Atik Budi Paryanti. (2022). PENGARUH KUALITAS LAYANAN DAN HARGA TERHADAP KEPUASAN PELANGGAN PADA GRABFOOD (STUDI WILAYAH KECAMATAN SETIABUDI). *JURNAL ILMIAH M-PROGRESS*, 12(1). <https://doi.org/10.35968/M-PU.V12I1.867>

Suryanto, A. A., Muqtadir, A., & Artikel, S. (2019). PENERAPAN METODE MEAN ABSOLUTE ERROR (MEA) DALAM ALGORITMA REGRESI LINEAR UNTUK PREDIKSI PRODUKSI PADI. *SAINTEKBU*, 11(1), 78–83. <https://doi.org/10.32764/SAINTEKBU.V11I1.298>

Usman, S., & Syam, R. F. (2024). Predictive Sparepart Maintenance Menggunakan Algoritma Machine Learning Extreme Gradient Boosting Regressor. *Journal of System and Computer Engineering*, 5(2), 249–258. <https://doi.org/10.61628/JSCE.V5I2.1418>

Wigati, A., Susanti, F., Program, M., Fakultas Ekonomi, S., Bisnis, D., & Terbuka, U. (2024). STUDI PENGARUH HARGA DAN PROMOSI TERHADAP KEPUTUSAN PEMBELIAN PRODUK FMCG (FAST MOVING CONSUMER GOODS) DI RANTAU KALIMANTAN SELATAN. *JURNAL AKADEMIK EKONOMI DAN MANAJEMEN*, 1(4), 186–192. <https://doi.org/10.61722/JAEM.V1I4.3320>

Winurputra, R., Ratnawati, D. E., & Korespondensi, P. (2025). Peramalan Penjualan Produk Menggunakan Extreme Gradient Boosting (XGBoost) dan Kerangka Kerja CRISP-DM untuk Pengoptimalan Manajemen Persediaan (Studi Kasus: UB Mart). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(2), 417–428. <https://doi.org/10.25126/JTIK.2025129451>

Yudiana, Y., Yulia Agustina, A., & Nur Khofifah, dan. (2023a). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1–20. <https://doi.org/10.30631/IJOIEB.V8I1.1710>

Yudiana, Y., Yulia Agustina, A., & Nur Khofifah, dan. (2023b). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1–20. <https://doi.org/10.30631/IJOIEB.V8I1.1710>

Yunitarini, R., Gultom, J. F., & Stefany, E. M. (2024). Klasifikasi Jamu Tradisional Madura Menggunakan Metode K-Nearest Neighbors (KNN) dan Term Frequency-Inverse Document Frequency (TF-IDF) Sebagai Representasi Teks. *Jurnal Informatika Polinema*, 11(1), 99–106. <https://doi.org/10.33795/JIP.V11I1.6456>

Zahrotul Kamalia, A., Nawangsih, I., & Korespondensi, P. (2025). Identifikasi Pola Tingkat Kesenjangan Ketuntasan Pendidikan Di Indonesia Dengan Menggunakan Metode K-Medoids Clustering. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(2), 321–330. <https://doi.org/10.25126/JTIK.2025129219>