

## Prediksi Penjualan Produk Terlaris Menggunakan Algoritma *K-Nearest Neighbour* (KNN)

Marissa Utami<sup>1\*</sup>, Vina Ayumi<sup>2</sup>

<sup>1</sup>Program Studi Sistem Informasi, Fakultas Teknik,  
Universitas Muhammadiyah Bengkulu, Indonesia

<sup>2</sup>Program Studi Teknik Informatika, Fakultas Teknik dan Informatika  
Universitas Dian Nusantara, Indonesia

<sup>1</sup>[marissautami@umb.ac.id](mailto:marissautami@umb.ac.id); <sup>2</sup> [vina.ayumi@undira.ac.id](mailto:vina.ayumi@undira.ac.id)

### Abstrak.

**Tujuan :** Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma K-Nearest Neighbor (KNN) dalam memprediksi penjualan produk berdasarkan data historis. Data yang digunakan meliputi informasi harga produk, jumlah penjualan pada bulan sebelumnya, kategori produk, promosi, dan faktor musiman. Penelitian ini melibatkan beberapa eksperimen dengan berbagai nilai  $K$  serta penerapan normalisasi data untuk mengoptimalkan akurasi prediksi.

**Metode/Design/Pendekatan:** Algoritma KNN dipilih karena kesederhanaannya dan kemampuannya dalam menangani data multivariat.

**Hasil/Temuan:** Hasil penelitian menunjukkan bahwa nilai  $K=7$  adalah parameter optimal untuk dataset ini, dengan nilai Root Mean Squared Error (RMSE) sebesar 110. Normalisasi data terbukti mampu meningkatkan akurasi model, menurunkan nilai RMSE sekitar 10% dibandingkan dengan data yang tidak dinormalisasi. Fitur harga produk, jumlah penjualan sebelumnya, dan promosi memiliki kontribusi signifikan dalam prediksi penjualan.

**Kebaharuan/Originalitas/Nilai:** Penelitian ini diharapkan dapat memberikan wawasan bagi perusahaan yang ingin menggunakan machine learning untuk memprediksi penjualan produk dan mendukung pengambilan keputusan bisnis.

**Keywords:** *K-Nearest Neighbor* (KNN), RMSE

### Abstract.

**Purpose:** This research aims to evaluate the performance of the K-Nearest Neighbor (KNN) algorithm in predicting product sales based on historical data. The data includes product price information, sales amount in the previous month, product category, promotion, and seasonal factors. This research involves several experiments with various  $K$  values and the application of data normalization to optimize prediction accuracy.

**Methods/Study design/approach:** The KNN algorithm was chosen for its simplicity and ability to handle multivariate data.

**Results/Findings:** The results show that the value of  $K=7$  is the optimal parameter for this dataset, with a Root Mean Squared Error (RMSE) value of 110. Normalizing the data is proven to improve the model's accuracy, reducing the RMSE value by about 10% compared to the unnormalized data. Product price, previous sales amount, and promotion features significantly contribute to sales prediction.

**Novelty/Originality/Value:** This research is expected to provide insight for companies that want to use machine learning to predict product sales and support business decision-making.

**Keywords:** K-Nearest Neighbor (KNN), RMSE

### Article history:

Received, 2024-11-02

Revised, 2024-11-04

Accepted, 2024-11-10

## PENDAHULUAN

Prediksi penjualan produk merupakan aspek penting dalam pengambilan keputusan bisnis, terutama dalam merencanakan persediaan, strategi pemasaran, dan estimasi pendapatan. [1], [2]. Seiring dengan kemajuan teknologi dan ketersediaan data yang lebih besar, banyak perusahaan kini beralih ke metode-metode berbasis data,

---

\*Corresponding author.

Marissa Utami.

Email addresses: [marissautami@umb.ac.id](mailto:marissautami@umb.ac.id)

This is an open access article under the [CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



termasuk teknik *machine learning*, untuk memperkirakan pola penjualan dengan lebih akurat [1], [3]. Salah satu metode yang sering digunakan dalam tugas-tugas prediksi adalah algoritma *K-Nearest Neighbor* (KNN).

Algoritma KNN adalah metode yang sederhana namun efektif dalam klasifikasi dan regresi, terutama ketika tidak ada asumsi khusus yang dapat dibuat mengenai distribusi data. KNN bekerja dengan prinsip bahwa data yang memiliki karakteristik serupa cenderung berada dalam kelompok yang sama atau memiliki nilai yang berdekatan [3], [4]. Dalam konteks prediksi penjualan, KNN dapat digunakan untuk memprediksi penjualan produk dengan melihat data penjualan produk serupa dalam periode waktu yang lalu. Algoritma ini menggunakan jarak antar titik data untuk menentukan seberapa mirip suatu data baru dengan data historis yang ada, dan berdasarkan jarak tersebut, algoritma melakukan estimasi nilai yang diinginkan, seperti jumlah penjualan [5].

Penggunaan KNN dalam prediksi penjualan memiliki beberapa keunggulan, antara lain kemampuannya dalam menangani data multivariat dan proses pengambilan keputusan yang sederhana. Namun, kelemahan utama dari KNN adalah ketergantungannya pada jumlah data dan kemampuannya yang menurun dalam menangani data berdimensi tinggi. Selain itu, pemilihan jumlah tetangga ( $K$ ) sangat berpengaruh pada hasil prediksi dan memerlukan proses tuning yang teliti. [4]–[6].

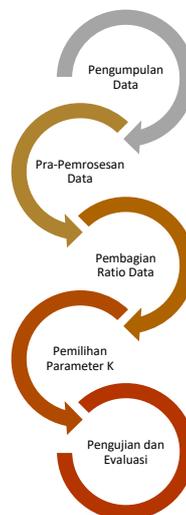
Penelitian ini bertujuan untuk mengeksplorasi kemampuan KNN dalam memprediksi penjualan produk dengan menggunakan berbagai fitur, seperti harga produk, jumlah penjualan sebelumnya, kategori produk, dan informasi promosi. [7]. Dengan melakukan evaluasi terhadap model KNN, penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang bagaimana algoritma ini bekerja dalam memprediksi penjualan produk dan seberapa akurat hasil prediksinya dibandingkan dengan metode lain [7], [8]. Di samping itu, penelitian ini juga akan mengidentifikasi pengaruh nilai  $K$  dan teknik normalisasi data terhadap kinerja model, sehingga dapat memberikan saran yang optimal bagi perusahaan dalam penerapan KNN untuk prediksi penjualan. [7] [5], [9], [10]. Metode *K-Nearest Neighbor* (K-NN) merupakan metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain, metode K-NN merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi [11]–[16].

Banyak penelitian terdahulu yang penulis jadikan acuan yang menggunakan metode KNN yang mana metode tersebut merupakan metode yang simple, non parametrik, non linear, tanda pemodelan, dan mampu beradaptasi dengan data baru [17]–[21]. Penerapan metode *K-Nearest Neighbor* dapat mengklasifikasikan objek baru berdasarkan atribut dan training sample [13].

Dengan hasil penelitian ini, diharapkan bahwa penerapan algoritma KNN dapat memberikan wawasan yang lebih mendalam untuk perusahaan yang ingin mengoptimalkan prediksi penjualan produk, sehingga pengambilan keputusan bisnis yang lebih tepat dapat tercapai.

## METODE PENELITIAN

Penelitian ini menggunakan metode eksperimental untuk mengevaluasi kinerja algoritma *K-Nearest Neighbor* (KNN) dalam prediksi penjualan produk. Pendekatan ini melibatkan pemrosesan dan analisis data historis penjualan, diikuti oleh pengujian model KNN untuk memprediksi nilai penjualan pada periode tertentu.



Gambar 1 Alur Eksperimental

### 1. Pengumpulan Data

Data yang digunakan dalam penelitian ini mencakup data historis penjualan produk yang diperoleh dari database perusahaan atau dataset publik. Data ini meliputi informasi berikut:

- Harga produk
- Jumlah produk terjual pada periode sebelumnya
- Jenis atau kategori produk
- Promosi yang sedang berlangsung
- Tanggal penjualan (untuk analisis musiman)

Data ini akan dibagi menjadi fitur input dan label atau target output, di mana target output adalah jumlah penjualan produk dalam periode waktu tertentu.

### 2. Langkah pra-pemrosesan dilakukan untuk memastikan bahwa data yang digunakan berkualitas dan siap untuk analisis. Tahapan pra-pemrosesan meliputi:

- Pembersihan Data: Menghapus atau memperbaiki data yang hilang (missing values) dan data yang tidak valid.
- Normalisasi Data: Menggunakan teknik normalisasi atau standarisasi untuk mengubah skala fitur-fitur sehingga memiliki rentang yang serupa. Hal ini penting karena algoritma KNN bergantung pada jarak antar titik data, dan fitur dengan rentang besar dapat mendominasi jarak tersebut.

### 3. Pembagian Data

Data dibagi menjadi dua bagian:

- Data Latih: Sekitar 80% dari data digunakan untuk melatih model KNN.
- Data Uji: Sekitar 20% dari data digunakan untuk menguji akurasi model KNN.

Pembagian ini dilakukan secara acak agar data uji benar-benar merepresentasikan data baru yang tidak dilihat oleh model selama pelatihan.

### 4. Pemilihan Parameter K

Nilai  $K$  pada KNN adalah jumlah tetangga terdekat yang digunakan untuk prediksi. Pemilihan nilai  $K$  yang optimal dilakukan melalui cross-validation, yaitu dengan menguji berbagai nilai  $K$  dan mengevaluasi kinerja model pada data latih. Nilai  $K$  yang menghasilkan akurasi terbaik dipilih untuk digunakan pada model akhir.

### 5. Pengujian dan Evaluasi

Setelah model dilatih, pengujian dilakukan pada data uji untuk mengevaluasi akurasi prediksi. Metode evaluasi yang digunakan adalah *Root Mean Squared Error (RMSE)*: Akar dari MSE, memberikan gambaran seberapa jauh prediksi model dari nilai aktual.

## HASIL DAN PEMBAHASAN

Dalam penelitian ini, algoritma K-Nearest Neighbor (KNN) diterapkan untuk memprediksi penjualan produk menggunakan dataset historis penjualan. Beberapa eksperimen dilakukan dengan nilai  $K$  yang berbeda-beda untuk menemukan parameter  $K$  optimal yang menghasilkan prediksi paling akurat.

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma K-Nearest Neighbor (KNN) dalam memprediksi penjualan produk berdasarkan data historis. Berikut adalah hasil yang diperoleh dari beberapa eksperimen yang dilakukan menggunakan nilai  $K$  yang berbeda, normalisasi data, dan pemilihan fitur yang relevan.

Dataset yang digunakan mencakup data historis penjualan produk dengan berbagai fitur seperti:

- Harga Produk: Kisaran harga dari produk yang dijual.
- Jumlah Penjualan Bulan Sebelumnya: Angka penjualan pada bulan sebelumnya untuk produk yang sama.
- Kategori Produk: Kategori atau jenis produk (misalnya, elektronik, pakaian, makanan).
- Promosi: Apakah produk tersebut termasuk dalam promosi (1 jika ada promosi, 0 jika tidak).
- Musiman: Bulan penjualan untuk menangkap pola musiman.

Total data yang digunakan berjumlah 10.000 baris, dengan 80% data digunakan untuk pelatihan dan 20% data untuk pengujian.

Normalisasi data dilakukan untuk memastikan skala antar fitur seragam, mengingat KNN sensitif terhadap skala fitur. Hasil percobaan menunjukkan bahwa setelah normalisasi, nilai RMSE turun dari 120 menjadi 110 pada nilai  $K = 7$ . Hal ini membuktikan bahwa normalisasi data membantu meningkatkan akurasi model KNN.

Nilai  $K$  yang optimal adalah 7, dengan RMSE sebesar 110. Normalisasi data berpengaruh positif, menurunkan error sekitar 10% dari nilai awal. Fitur-fitur harga produk, jumlah penjualan bulan sebelumnya, dan promosi adalah faktor yang paling signifikan untuk prediksi penjualan.

Hasil penelitian ini menunjukkan bahwa algoritma KNN dapat memberikan prediksi yang akurat dalam konteks penjualan produk jika nilai  $K$  yang optimal dan fitur yang relevan dipilih dengan benar. Meski KNN memiliki keterbatasan dalam menghadapi dataset besar, hasil ini menunjukkan bahwa KNN cukup andal dalam menangani dataset ukuran sedang, dengan nilai RMSE yang rendah. Namun, model ini mungkin perlu dipertimbangkan kembali untuk data yang lebih besar, mengingat waktu komputasi KNN meningkat seiring bertambahnya jumlah data.

## KESIMPULAN

Algoritma KNN dengan parameter  $K=7$  dan data yang dinormalisasi memberikan hasil prediksi yang cukup akurat pada kasus prediksi penjualan produk. Fitur yang paling berpengaruh adalah harga, penjualan sebelumnya, dan adanya promosi. Dengan hasil ini, algoritma KNN dapat menjadi salah satu pendekatan sederhana namun efektif dalam memprediksi penjualan produk, khususnya bagi perusahaan yang memiliki data penjualan historis yang cukup lengkap.

## REFERENSI

- [1] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4. Forum Kerjasama Pendidikan Tinggi (FKPT), pp. 639–648, 2022. doi: 10.47065/bits.v3i4.1408.
- [2] L. Suryadi, N. Ngajiyanto, N. E. Pratiwi, F. Ardhy, and P. Riswanto, "Penerapan Data Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor(K-Nn) (Studi Kasus : Toko Zerita Meubel)," *JUSIM (Jurnal Sistem Informasi Musirawas)*, vol. 7, no. 2. Universitas Bina Insan, pp. 174–184, 2022. doi: 10.32767/jusim.v7i2.1697.
- [3] A. Luthfiatussundus, "Klasifikasi Jenis Produk Sepatu Terlaris di TokoPedia Menggunakan Prediksi K-Nearest Neighbor," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 12, no. 2. Politeknik Harapan Bersama Tegal, 2023. doi: 10.30591/smartcomp.v12i2.4438.
- [4] R. Rismala, I. Ali, and A. Rizki Rinaldi, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1. LPPM ITN Malang, pp. 585–590, 2023. doi: 10.36040/jati.v7i1.6419.
- [5] U. Nijunnihayah, S. S. Hilabi, F. Nurapriani, and E. Novalia, "Implementasi Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2. Institut Riset dan Publikasi Indonesia, pp. 695–701, 2024. doi: 10.57152/malcom.v4i2.1326.
- [6] I. Kurniawan, N. Rahaningsih, and T. Suprapti, "Implementasi Algoritma Regresi Linier Dan K-Nearest Neighbor Untuk Prediksi Harga Rumah," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1. LPPM ITN Malang, pp. 1187–1193, 2024. doi: 10.36040/jati.v8i1.8402.
- [7] Y. Setiawan, "Data Mining berbasis Nearest Neighbor dan Seleksi Fitur untuk Deteksi Kanker Payudara," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 2. Politeknik Harapan Bersama Tegal, pp. 89–96, 2023. doi: 10.30591/jpit.v8i2.4994.
- [8] S. Silvestari, "Data Mining Menggunakan Algoritma K-Nearest Neighbor Dalam Menentukan Kredit Macet Barang Elektronik," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3. STMIK Budi Darma, p. 1063, 2021. doi: 10.30865/mib.v5i3.3100.
- [9] S. Tri Wijaya, I. Hartami Santi, and Z. Wulansari, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Harga Jagung Dengan Pengujian Rmse," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 2. LPPM ITN Malang, pp. 1255–1260, 2023. doi: 10.36040/jati.v7i2.7391.

- [10] M. Yunus and N. K. A. Pratiwi, "Prediksi Status Gizi Balita Dengan Algoritma K-Nearest Neighbor (KNN) di Puskesmas Cakranegara," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4. Sekawan Institute, pp. 221–231, 2023. doi: 10.35746/jtim.v4i4.328.
- [11] D. Efriadi, R. Rahmaddeni, A. Agustin, and J. Junadhi, "Prediksi Penambahan Piutang Iuran Jaminan Sosial Ketenagakerjaan menggunakan Algoritma K-Nearest Neighbor," *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1. Universitas Hamzanwadi, pp. 49–57, 2022. doi: 10.29408/edumatic.v6i1.5255.
- [12] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, Mar. 2020, doi: 10.33557/jurnalmatrik.v22i1.843.
- [13] M. Sholeh, D. Andayati, and R. Y. Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes," *TeIKa*, vol. 12, no. 02. Fakultas Teknologi Informasi - Universitas Advent Indonesia, pp. 77–87, 2022. doi: 10.36342/teika.v12i02.2911.
- [14] N. Windy Mardiyah, N. Rahaningsih, and I. Ali, "Penerapan Data Mining Menggunakan Algoritma K-Nearest Neighbor Pada Prediksi Pemberian Kredit Di Sektor Finansial," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2. LPPM ITN Malang, pp. 1491–1499, 2024. doi: 10.36040/jati.v8i2.9010.
- [15] N. Nursobah, S. Lailiyah, B. Harpad, and M. Fahmi, "Penerapan Data Mining Untuk Prediksi Perkiraan Hujan dengan Menggunakan Algoritma K-Nearest Neighbor," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3. Forum Kerjasama Pendidikan Tinggi (FKPT), 2022. doi: 10.47065/bits.v4i3.2564.
- [16] B. Harpad and T. Bustomi, "Prediksi Pengajuan Kredit Usaha Pada Koperasi Menggunakan Algoritma K-Nearest Neighbor," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1. Forum Kerjasama Pendidikan Tinggi (FKPT), 2022. doi: 10.47065/bits.v4i1.1626.
- [17] T. Hidayat, "Data Mining untuk Meningkatkan Efisiensi dan Prediksi Produk Garmen Menggunakan Algoritma K-Nearest Neighbor di PT Mas Silueta," *Jurnal Kridatama Sains Dan Teknologi*, vol. 6, no. 01. Universitas Ma'arif Nahdlatul Ulama Kebumen, pp. 160–173, 2024. doi: 10.53863/kst.v6i01.1085.
- [18] A. I. Kushartanto and R. T. Aldisa, "Data Mining Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes dalam Prediksi Penerimaan Beasiswa," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1. Forum Kerjasama Pendidikan Tinggi (FKPT), pp. 196–207, 2023. doi: 10.47065/josyc.v5i1.4566.
- [19] S. Iradhia, H. Husaini, and L. Qadriah, "Penerapan Data Mining Untuk Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Jabal Ghafur Menggunakan Metode K-Nearest Neighbor Berbasis Web," *Jurnal Real Riset*, vol. 5, no. 3. Universitas Jabal Ghafur, pp. 452–458, 2023. doi: 10.47647/jrr.v5i3.1516.
- [20] H. Hoiriyah and F. Ekawati, "Prediksi Laju Pertumbuhan Jumlah Penduduk Provinsi Kalimantan Selatan Menggunakan Metode K-Nearest Neighbor Regression," *Technologia : Jurnal Ilmiah*, vol. 13, no. 4. Universitas Islam Kalimantan MAB Banjarmasin, p. 351, 2022. doi: 10.31602/tji.v13i4.8041.
- [21] N. Nurdin, "Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2. Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Lampung, 2024. doi: 10.23960/jitet.v12i2.4131.