



Prediksi Penjualan Barang Menggunakan Algoritma K-Nearest Neighbor

Filbert Wijaya¹, Rizky Rahmansyah², Nelly Astuti Hasibuan³

¹Teknik Informatika, Universitas Prima Indonesia, Medan, Indonesia, filbert.wijaya.004@gmail.com

²Teknik Informatika, Universitas Prima Indonesia, Medan, Indonesia, rizkyrahmans2021@gmail.com

³Teknik Informatika, Universitas Prima Indonesia, Medan, Indonesia, nelly.ahsb@gmail.com

STATUS ARTIKEL

Dikirim 10 Agustus 2025

Direvisi 10 September 2025

Diterima 31 Oktober 2025

Kata Kunci:

K-Nearest Neighbor, Data Mining, Prediksi, Mean Absolute Percentage Error, Mean Absolute Error

ABSTRAK

Kemampuan untuk memprediksi jumlah penjualan produk sangat penting dalam mendukung pengambilan keputusan bisnis seperti pengelolaan stok dan perencanaan produksi. Penelitian ini bertujuan membangun model prediksi penjualan menggunakan algoritma K-Nearest Neighbor (KNN) berdasarkan data historis penjualan enam jenis produk dari Toko X pada periode Agustus 2022 hingga November 2023. Proses prediksi melibatkan penentuan nilai k , perhitungan jarak, pemilihan tetangga terdekat, dan evaluasi menggunakan metrik Mean Absolute Percentage Error (MAPE) dan Mean Absolute Error (MAE). Hasil evaluasi menunjukkan bahwa model KNN mampu memprediksi tren penjualan dengan akurasi yang cukup baik, dengan nilai rata-rata MAPE sebesar 16,66% dan MAE sebesar 4168,77. Model ini dapat digunakan sebagai alat bantu untuk memperkirakan permintaan barang dan mendukung pengambilan keputusan operasional secara lebih efektif dan efisien.

1. PENDAHULUAN

Kemampuan untuk memprediksi jumlah produk yang akan dijual sangat penting dalam membuat keputusan terkait strategi seperti manajemen stok, perencanaan produksi, dan strategi pemasaran di dunia bisnis dan perdagangan. Ketidaksesuaian dalam memprediksi permintaan barang dapat menyebabkan kelebihan atau kekurangan barang, sehingga akan berdampak pada kepuasan pelanggan dan juga kerugian dalam finansial. Dalam dunia bisnis dan perdagangan, diperlukan sebuah sistem yang dapat menggunakan data historis guna memprediksi penjualan secara tepat dan akurat.

Pendekatan konvensional dalam melakukan analisis data penjualan menjadi tidak efektif dikarenakan jumlah data yang terus meningkat. Oleh karena itu, pendekatan berbasis kecerdasan buatan seperti machine learning mulai populer. K-Nearest Neighbor (KNN) adalah yang paling banyak digunakan untuk membuat model prediksi berbasis data. Algoritma ini dianggap sederhana namun efektif, terutama untuk menemukan pola berdasarkan kedekatan antar data.

Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi penjualan berbasis KNN yang dapat membantu bisnis membuat keputusan yang lebih baik tentang perencanaan produksi dan pengelolaan stok. Dengan prediksi yang lebih akurat, bisnis dapat menghasilkan lebih banyak uang dan lebih banyak kepuasan pelanggan.

Sebuah toko yang menjual berbagai jenis kertas bernama Toko X akan menjadi tempat data penjualan dikumpulkan. Jenis kertas yang dijual pun beragam, tetapi di sini akan dikelompokkan menjadi enam jenis kertas, yaitu CraftFoodPack, Dupleks, FoodPackMatte, GreaseProof, HVS, dan Ivory. Data diambil dari bulan Agustus 2022 sampai dengan November 2023.

2. METODE

2.1 Tahapan Penelitian

Tahapan penelitian ini dapat dilihat dari gambar berikut.



Gambar 2.1 Tahapan Penelitian

Berikut ini adalah penjelasan tahapan penelitian di atas:

- Identifikasi masalah dilakukan untuk menentukan isu utama yang relevan dan layak dijadikan sebagai fokus penelitian.
- Studi literatur merupakan proses mengumpulkan dan menganalisis informasi dari berbagai sumber yang relevan, seperti jurnal, artikel ilmiah, dan laporan penelitian.
- Pengumpulan data dilakukan untuk memperoleh informasi dalam mencari solusi permasalahan.
- Analisis data merupakan proses untuk memperoleh informasi yang dapat dijadikan dasar pengambilan keputusan. Proses ini meliputi pengelompokan, pembersihan, transformasi, dan pemodelan data.
- Perancangan dan pembangunan sistem bertujuan menciptakan atau menyempurnakan program untuk memenuhi kebutuhan pengguna.
- Uji coba sistem merupakan proses pengujian program yang sudah dibuat untuk menilai apakah program tersebut berjalan dengan semestinya atau tidak.
- Implementasi sistem merupakan tahap di mana rancangan sistem yang telah disusun, direalisasikan menjadi sebuah aplikasi atau program yang berfungsi secara nyata.

2.2 Analisis Data

Analisis data adalah proses pengolahan data untuk memperoleh informasi yang berguna dan dapat digunakan sebagai solusi suatu permasalahan. Berikut ini adalah data penjualan seluruh produk Toko X dari Agustus 2022 sampai November 2023.

Tabel 2.1 Jumlah Penjualan Setiap Produk

No	Jenis Produk	Jumlah Terjual
----	--------------	----------------

1	CraftFoodpack	403000
2	Dupleks	472500
3	FoodpackMatte	232500
4	GreaseProof	137500
5	HVS	218500
6	Ivory	517500

2.3 Data Mining

Data mining adalah proses pengumpulan dan pengolahan data [4] yang bertujuan untuk mengekstrak informasi penting pada data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi Artificial Intelligence (AI). Data mining memiliki beberapa manfaat, seperti:

- Meningkatkan Efisiensi Operasional. Dengan data mining, perusahaan dapat mengidentifikasi bagian dari proses bisnis yang kurang efisien, serta menemukan peluang untuk melakukan perbaikan.
- Memprediksi Tren Pasar. Melalui analisis data historis, data mining dapat membantu memperkirakan tren pasar di masa mendatang, sehingga perusahaan dapat menyesuaikan strategi mereka sesuai dengan kebutuhan pasar.

2.4 Machine Learning

Machine learning merupakan cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) yang berfokus pada pengembangan sistem yang mampu belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit. Dalam konteks ini, machine learning digunakan untuk membangun model prediktif dari data historis penjualan agar dapat memperkirakan jumlah penjualan di waktu mendatang.

Dengan menggunakan algoritma machine learning seperti KNN, model dapat mempelajari pola dari data penjualan yang telah ada dan menggunakannya untuk melakukan prediksi secara otomatis dan adaptif.

2.5 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan salah satu algoritma dalam data mining yang termasuk ke dalam metode klasifikasi berbasis instance (*lazy learning*). Berbeda dengan algoritma lain yang membentuk model saat proses pelatihan, KNN tidak membentuk model secara eksplisit. Sebaliknya, KNN menyimpan seluruh data latih dan melakukan proses klasifikasi atau prediksi hanya saat menerima data baru.

KNN bekerja dengan cara menghitung jarak antara data yang ingin diprediksi dengan seluruh data latih, kemudian mengambil sejumlah k tetangga terdekat (*nearest neighbors*). Nilai k adalah jumlah tetangga terdekat yang akan digunakan untuk menentukan kelas atau nilai dari data baru. Kelas atau nilai hasil prediksi ditentukan berdasarkan mayoritas kelas dari tetangga tersebut (dalam klasifikasi) atau rata-rata nilai (dalam regresi).

3. HASIL DAN PEMBAHASAN

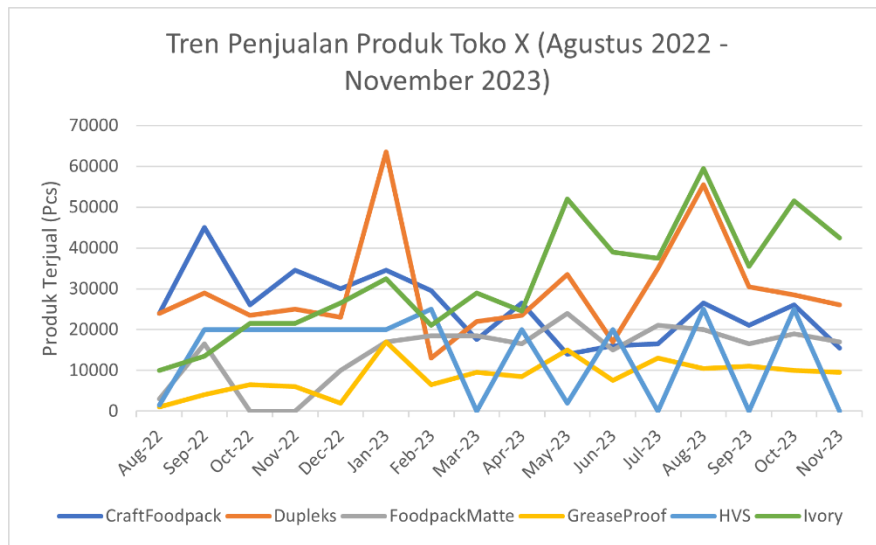
3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini merupakan data penjualan dari Toko X yang mencakup enam jenis produk, yaitu: CraftFoodpack, Dupleks, FoodpackMatte, GreaseProof, HVS, dan Ivory. Data dicatat secara bulanan dari Agustus 2022 hingga November 2023. Setiap entri data terdiri dari tiga atribut utama:

- Bulan : periode penjualan dalam format MM/YYYY.
- Jenis Produk : kategori produk yang dijual.
- Jumlah Order : jumlah unit produk yang terjual pada bulan tersebut.

3.2 Visuliasasi Data

Untuk memahami perilaku penjualan setiap produk, dilakukan visualisasi tren penjualan berdasarkan bulan. Berikut ini adalah grafik penjualan setiap produk:



Gambar 3.1 Diagram Garis Penjualan Barang

- CraftFoodpack: fluktuatif, puncak penjualan pada November 2022 dan Januari 2023, penurunan tajam pada November 2023.
- Dupleks: mengalami kenaikan signifikan di Januari dan Agustus 2023.
- FoodpackMatte: mengalami kenaikan signifikan di Mei 2023 dan cenderung stabil di tahun 2023.
- GreaseProof: mengalami kenaikan signifikan di Januari 2023.
- HVS: tren tidak konsisten, beberapa bulan tidak ada penjualan.
- Ivory: relatif stabil dan cenderung meningkat, mencapai puncak pada Agustus 2023.

3.3 Proses Prediksi Penjualan Menggunakan K-Nearest Neighbor (KNN)

Setelah data historis penjualan dianalisis dan divisualisasikan, langkah selanjutnya adalah membangun model prediksi menggunakan algoritma K-Nearest Neighbor (KNN). Proses ini meliputi beberapa tahapan utama:

- Penentuan nilai k. Nilai k (jumlah tetangga terdekat) dipilih melalui uji coba dan teknik validasi silang (cross-validation). Pemilihan nilai k yang optimal berfungsi untuk mendapatkan keseimbangan terbaik antara bias dan varians, sehingga model bisa menghasilkan prediksi yang akurat dan tidak overfitting atau underfitting. Dalam penelitian ini, nilai k yang digunakan adalah 3.
- Penghitungan jarak. Data penjualan yang terdiri dari variabel seperti bulan, jenis produk, dan jumlah terjual, digunakan untuk menghitung jarak antara data baru dengan data historis menggunakan metrik jarak.
- Pemilihan tetangga terdekat. Berdasarkan perhitungan jarak, diambil k data terdekat yang akan digunakan sebagai referensi prediksi.
- Prediksi nilai penjualan. Nilai penjualan yang diprediksi didasarkan pada rata-rata jumlah penjualan dari k tetangga terdekat tersebut.

- Evaluasi model. Kinerja model diukur menggunakan metrik regresi, seperti Mean Absolute Error (MAE) dan Mean Absolute Percentage Error (MAPE), untuk menilai seberapa dekat prediksi dengan data aktual.

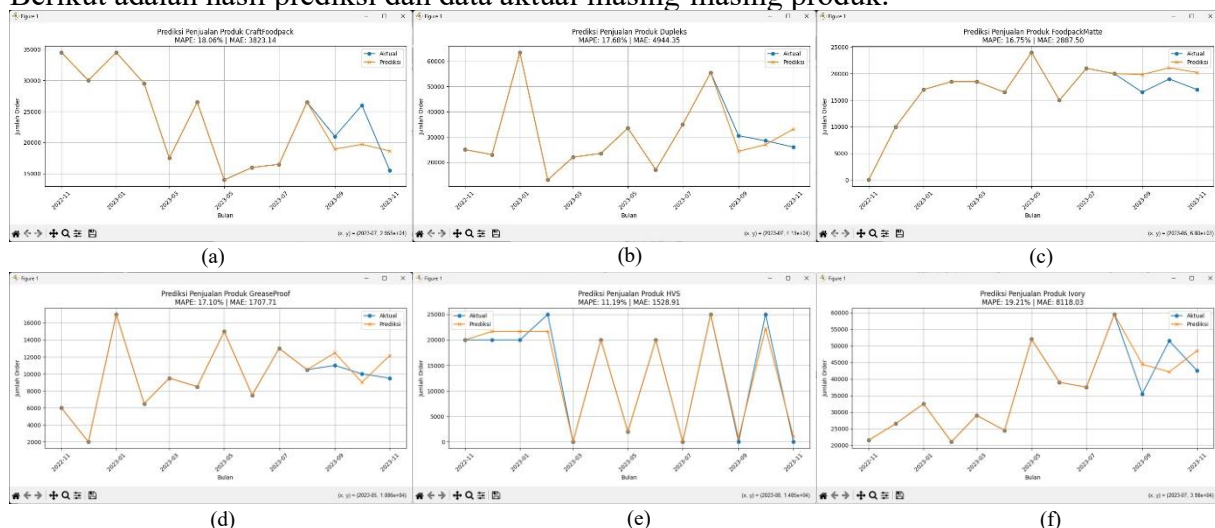
3.4 Visualisasi dan Evaluasi Hasil Prediksi

Setelah model K-Nearest Neighbor (KNN) dilatih dan diuji, hasil prediksi penjualan untuk masing-masing produk dibandingkan dengan data penjualan aktual. Perbandingan ini divisualisasikan dalam grafik yang menampilkan kurva penjualan aktual dan kurva hasil prediksi sepanjang periode pengamatan.

Setiap grafik menunjukkan seberapa baik model KNN mampu menangkap pola penjualan berdasarkan data historis. Untuk memberikan gambaran kuantitatif tentang keakuratan model, nilai Mean Absolute Percentage Error (MAPE) dan Mean Absolute Error (MAE) turut disertakan pada setiap grafik. MAPE dan MAE yang kecil menunjukkan performa model yang lebih baik dan prediksi yang lebih akurat.

- MAPE mengukur rata-rata persentase kesalahan absolut antara nilai prediksi dan nilai aktual, memberikan indikasi seberapa besar kesalahan prediksi secara relatif terhadap nilai aktual, biasanya dinyatakan dalam persen.
- MAE mengukur rata-rata nilai absolut selisih antara prediksi dan realisasi, yang menunjukkan besarnya kesalahan rata-rata tanpa memperhatikan arah kesalahan.

Berikut adalah hasil prediksi dan data aktual masing-masing produk.



Gambar 3.2 Penjualan Aktual vs Prediksi (a) CraftFoodPack, (b) Dupleks, (c) FoodPackMatte, (d) GreaseProof, (e) HVS, (f) Ivory

Penjelasan:

- CraftFoodPack: Model berhasil mengikuti tren utama penjualan dengan baik. Nilai MAPE sebesar 18,06% menunjukkan tingkat akurasi yang cukup baik.
- Dupleks: Prediksi model cukup sesuai dengan pola aktual. MAPE sebesar 17,68% menunjukkan tingkat akurasi yang cukup baik.
- FoodPackMatte: Model mampu menangkap fluktuasi ringan dalam data aktual. Nilai MAPE sebesar 16,75% menunjukkan prediksi yang cukup akurat.
- GreaseProof: Prediksi model cukup sesuai dengan pola aktual. MAPE sebesar 17,10% menunjukkan tingkat akurasi yang cukup baik.

- HVS: Prediksi cukup sulit karena data aktual tidak konsisten dan beberapa bulan kosong. Walaupun nilai MAPE hanya 11,19%, prediksi bisa saja meleset dikarenakan MAPE tidak memasukkan nilai 0 pada data saat melakukan perhitungan.
- Ivory: Model memprediksi tren kenaikan dengan baik, sesuai dengan lonjakan aktual di pertengahan 2023. MAPE sebesar 19,21% menunjukkan kinerja prediksi yang cukup baik.

```

Rata-rata MAPE: 16.66%
Rata-rata MAE: 3834.94

Prediksi Penjualan Bulan Depan:
• CraftFoodpack: 18257 unit
• Dupleks: 34357 unit
• FoodpackMatte: 19969 unit
• GreaseProof: 11405 unit
• HVS: 22203 unit

```

Gambar 3.3 Hasil Program

Gambar di atas adalah hasil program yang menunjukkan nilai rata-rata MAPE adalah 16,66%. Hal ini berarti program dapat memprediksi penjualan dengan cukup baik.

Berikut adalah tabel yang menunjukkan nilai MAPE, MAE dan prediksi bulan depan masing-masing produk.

Tabel 3. 1 Nilai MAPE, MAE dan Prediksi Penjualan

No	Jenis Produk	MAPE	MAE	Prediksi Bulan Depan
1	CraftFoodpack	18,06%	3823,14	18257
2	Dupleks	17,68%	4944,35	34357
3	FoodpackMatte	16,75%	2887,50	19969
4	GreaseProof	17,10%	1707,71	11405
5	HVS	11,19%	1528,91	22203
6	Ivory	19,21%	8118,03	43971

4. KESIMPULAN

Data penjualan enam jenis produk di Toko X yang dianalisis menggunakan K-Nearest Neighbor (KNN) mencakup periode dari Agustus 2022 hingga November 2023, dengan atribut utama berupa bulan, jenis produk, dan jumlah order. Dari hasil analisis dan penerapan model, diperoleh beberapa kesimpulan utama sebagai berikut:

- Visualisasi menunjukkan pola penjualan yang bervariasi, dari fluktuatif hingga stabil, termasuk lonjakan pada bulan tertentu. Hal ini mengindikasikan adanya pengaruh musiman atau faktor eksternal terhadap penjualan.
- Algoritma KNN terbukti mampu memodelkan data penjualan dengan cukup baik, terutama untuk produk yang memiliki pola penjualan yang stabil dan konsisten.
- Evaluasi model menggunakan Mean Absolute Percentage Error (MAPE) dan Mean Absolute Error (MAE) menunjukkan bahwa nilai kesalahan prediksi masih dalam batas yang dapat diterima. Nilai rata-rata MAPE adalah 16,66% dan MAE adalah 4168,77.
- Prediksi penjualan bulan berikutnya memberikan estimasi yang dapat digunakan oleh Toko X sebagai acuan dalam pengelolaan stok dan strategi pemasaran. Produk dengan estimasi penjualan tertinggi adalah Ivory, sedangkan produk dengan estimasi terendah adalah GreaseProof.

5. DAFTAR PUSTAKA

- Alfani, A. W. P. R., Rozi, F., & Sukmana, F. (2021). Prediksi Penjualan Produk Unilever menggunakan Metode K-Nearest Neighbor. *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 6(1), 155–160. <https://doi.org/10.29100/jipi.v6i1.1910>
- Amanda, P. Y., Damayanti, B. A., Choirun, A. A., Sari, S. N., Armiyanti, S., & Hidayat, M. M. (2025). Penerapan Data Mining untuk Prediksi Penjualan Produk Skincare Menggunakan Metode KNN (K-Nearest Neighbors): Studi Kasus Klinik Ayume Beauty Care. *Dike*, 3(1), 25-30. <https://doi.org/10.69688/dike.v3i1.122>
- Anggrawan, A., Hairani, H., & Azmi, N. (2022). Prediksi Penjualan Produk Unilever menggunakan Metode Regresi Linear. *Jurnal Bumigora Information Technology (BITE)*, 4(2), 123-132. <https://doi.org/10.30812/bite.v4i2.2416>
- Astri, J., Karman, J., & Daulay, N. K. (2023). Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neighbor (KNN) pada Fakultas Ilmu Teknik, Univeritas Bina Insan. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 8(1), 169-173.
- Dewi, S. P., Nurwati, & Rahayu, E. (2022). Penerapan Data Mining untuk Prediksi Penjualan Produk Terlaris menggunakan Metode K-Nearest Neighbor. *Building Informatics Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- Maskuri, M. N., Harliana, Sukerti, K., & Bhakti, R. M. H. (2022). Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Prediksi Penyakit Stroke. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 4(1), 130–140. <https://doi.org/10.46772/intech.v4i01.751>
- 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.
- Rismayadi, A. A., Febrianto, R. W., Raharja, A. R., & Hariyanti, I. (2024). Perbandingan Kinerja Metode Machine Learning Support Vector Machine (SVM), Random Forest, dan K-Nearest Neighbors (KNN) dalam Prediksi Harga Saham Apple. *Media Informatika*, 23(3), 152-160. <https://doi.org/10.37595/mediainfo.v23i3.299>
- Simbolon, I. N., Siburian, H. D. S. J., & Man, W. A. (2023). Prediksi Kualitas Air Sungai di Jakarta menggunakan KNN yang Dioptimalisasi dengan PSO. *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, 12(2), 1193–1203. <https://doi.org/10.23960/jitet.v12i2.4191>
- Wafir, A. M., & Fatah, Z. (2025). Prediksi Produk Penjualan Di Supermarket Dengan Metode Algoritma K-Nearest Neighbors (KNN). *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, 3(1), 75-82. <https://doi.org/10.59024/jiti.v3i1.1056>