



PREDIKSI PRODUK PENJUALAN DI SUPERMARKET DENGAN METODE ALGORITMA K-NEAREST NEIGHBORS (KNN)

Ahmad Muflih Wafir S.A¹, Zaehol Fatah²

^{1,2} Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Jl. KHR. Syamsul Arifin, No.1-2, Situbondo 68374, Indonesia

muflih438@gmail.com^{1*}, zaeholfatah@gmail.com²

Abstract. *In today's era, there are already many companies that have been established, from urban to rural areas, various companies have been established, especially companies that provide daily necessities such as supermarkets. And each company competes with each other in selling its products with the expected results. In this study, researchers use data to support this study. Because sales of goods or product stock can be calculated in sales results, the higher the sales, the higher the risk that will be faced. This study aims to apply data mining in analyzing sales results that occur in supermarkets. And to find out its impact on sales. This researcher uses the KNN method, by looking for test results with this method which will be implemented using the Rapid Miner application which will later produce the results of its analysis.*

Keywords: *Data Mining, K-Nearest Neighbors, Sales*

Abstrak. Pada era sekarang ini sudah banyak sekali perusahaan yang berdiri, mulai dari perkotaan hingga pedesaan, berbagai macam perusahaan telah berdiri terutama perusahaan yang menyediakan kebutuhan sehari-hari seperti halnya supermarket. Dan setiap perusahaan tersebut saling bersaing dalam menjual produknya dengan hasil yang diharapkan. Pada penelitian ini peneliti menggunakan data untuk mendukung penelitian ini. Karena penjualan barang atau stok produk dapat dihitung dalam hasil penjualan, semakin tinggi penjualan maka semakin tinggi pula resiko yang akan dihadapi. Penelitian ini bertujuan untuk menerapkan data mining dalam menganalisis hasil penjualan yang terjadi pada supermarket. Dan untuk mengetahui dampaknya terhadap penjualan. Peneliti ini menggunakan metode KNN, dengan cara mencari hasil pengujian dengan metode ini yang akan diimplementasikan dengan menggunakan aplikasi Rapid Miner yang nantinya akan menghasilkan hasil analisisnya.

Kata kunci: Data Mining, K-Nearest Neighbors, Penjualan

1. LATAR BELAKANG

Dalam era digitalisasi saat ini, volume data yang dihasilkan dari transaksi penjualan di supermarket semakin meningkat secara signifikan. Data ini mencakup informasi mengenai jenis produk, jumlah penjualan, kategori produk, waktu transaksi, dan perilaku pelanggan. Namun, data tersebut sering kali belum dimanfaatkan secara optimal untuk membantu pengambilan keputusan yang lebih baik. Salah satu pendekatan yang dapat digunakan untuk menggali informasi berharga dari data tersebut adalah dengan menerapkan metode data mining. Data mining memungkinkan analisis data yang kompleks untuk menemukan pola-pola tersembunyi, salah satunya adalah memprediksi produk dengan penjualan tinggi.

Received: Desember, 2024; Revised Desember 4, 2024; Januari 2, 2025

*Ahmad Muflih Wafir S.A, muflih438@gmail.com

Penelitian ini berfokus pada penerapan algoritma K-Nearest Neighbors (KNN) untuk memprediksi produk dengan tingkat penjualan tertinggi di supermarket. KNN merupakan salah satu algoritma klasifikasi dalam data mining yang sederhana namun efektif. Prinsip kerja KNN adalah dengan mengukur jarak antara satu data dengan data lainnya menggunakan Euclidean Distance atau metode lainnya, untuk kemudian menentukan kategori berdasarkan tetangga terdekatnya. Dalam konteks penjualan supermarket, algoritma KNN dapat digunakan untuk memprediksi produk mana yang akan memiliki penjualan tinggi berdasarkan data historis penjualan.

Manajemen supermarket sering dihadapkan pada tantangan untuk mengatur stok produk agar sesuai dengan permintaan pelanggan. Ketidaktepatan dalam memprediksi penjualan dapat mengakibatkan kerugian, seperti kekurangan stok produk yang diminati atau kelebihan stok pada produk yang kurang laku. Oleh karena itu, prediksi yang akurat mengenai produk dengan penjualan tinggi dapat membantu pihak manajemen dalam mengambil keputusan yang lebih efisien terkait pengelolaan stok dan strategi pemasaran. Dengan menerapkan algoritma KNN, penelitian ini diharapkan dapat memberikan solusi sederhana namun akurat untuk masalah prediksi penjualan di supermarket.

2. KAJIAN TEORITIS

Teori Data Mining

Data mining adalah proses penemuan pola dan pengetahuan dari kumpulan data besar menggunakan metode statistik, algoritma, dan pembelajaran mesin. Dalam konteks penjualan di supermarket, data mining digunakan untuk mengidentifikasi pola perilaku konsumen, tren pembelian, serta prediksi permintaan produk.

Algoritma K-Nearest Neighbors

Algoritma K-Nearest Neighbors (KNN) adalah salah satu algoritma berbasis instance-based learning yang digunakan untuk klasifikasi atau regresi. Prinsip utama KNN adalah menentukan jarak terdekat antara data yang baru dengan data-data dalam dataset yang sudah diketahui kelasnya. Metode ini sering digunakan dalam analisis prediksi karena keunggulannya dalam menangani data non-linear dan sederhana dalam implementasi.

Rumus yang digunakan untuk menghitung jarak biasanya adalah Euclidean Distance, yang dirumuskan sebagai berikut:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Dengan p dan q adalah dua titik data dalam ruang dimensi n .

Bagian ini menguraikan teori-teori relevan yang mendasari topik penelitian dan memberikan ulasan tentang beberapa penelitian sebelumnya yang relevan dan memberikan acuan serta landasan bagi penelitian ini dilakukan. Jika ada hipotesis, bisa dinyatakan tidak tersurat dan tidak harus dalam kalimat tanya.

Penelitian Sebelumnya

Beberapa penelitian relevan yang menjadi acuan bagi studi ini adalah sebagai berikut:

- a) Sari dan Nugroho (2020) yang menggunakan KNN untuk memprediksi permintaan produk di toko ritel, menunjukkan akurasi sebesar 85%.
- b) Putra et al. (2021) yang menerapkan KNN untuk segmentasi pelanggan dalam rangka meningkatkan strategi pemasaran.
- c) Widodo dan Lestari (2023) yang membandingkan metode KNN dengan Decision Tree dalam prediksi penjualan, menunjukkan keunggulan KNN pada dataset dengan atribut numerik tinggi.

Kajian ini menjadi dasar teori dalam penelitian untuk mengimplementasikan KNN pada prediksi penjualan produk di supermarket, guna memberikan panduan strategi pengelolaan stok yang lebih optimal.

3. METODE PENELITIAN

Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode prediktif untuk memodelkan pola penjualan produk di supermarket. Data yang digunakan dalam penelitian ini merupakan data penjualan historis yang mencakup atribut seperti tanggal transaksi, jenis produk, jumlah penjualan, harga, dan kategori produk. Penelitian ini bertujuan untuk mengidentifikasi pola penjualan dan memprediksi produk yang berpotensi memiliki volume penjualan tinggi menggunakan algoritma K-Nearest Neighbors (KNN).

Teknik dan Instrumen Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari basis data penjualan supermarket. Data tersebut mencakup informasi berikut:

- a) Tanggal transaksi: untuk menganalisis pola musiman atau waktu tertentu.
- b) Jenis produk: untuk mengidentifikasi karakteristik produk.
- c) Jumlah penjualan: sebagai variabel target prediksi.
- d) Harga produk: untuk analisis korelasi harga dan penjualan.
- e) Kategori produk: untuk klasifikasi produk berdasarkan kelompok tertentu.

Model Penelitian

Model penelitian dirancang dengan menggunakan algoritma KNN. Algoritma ini bekerja berdasarkan penghitungan jarak Euclidean untuk mengidentifikasi data tetangga terdekat (nearest neighbors). Parameter utama dalam algoritma KNN adalah nilai k (jumlah tetangga) yang ditentukan melalui uji coba untuk mendapatkan kinerja terbaik.

4. HASIL DAN PEMBAHASAN

Proses Pengumpulan Data

Penelitian ini dilakukan dengan mengumpulkan data penjualan dari salah satu supermarket di Bali. Data yang digunakan meliputi atribut seperti tanggal transaksi, kategori produk, jumlah terjual, harga per unit, serta diskon yang diberikan. Lokasi penelitian berpusat pada supermarket besar dengan kategori produk makanan, minuman, dan kebutuhan rumah tangga.

Data yang diperoleh dari supermarket kemudian melalui proses preprocessing yang meliputi pembersihan data, pengisian nilai kosong, normalisasi data, dan pembagian data menjadi data latih (80%) dan data uji (20%).

Hasil Analisis Data

1. Penentuan Nilai K Optimal

Penentuan nilai optimal untuk parameter K dalam algoritma KNN dilakukan menggunakan teknik *cross-validation*. Dari pengujian beberapa nilai K (mulai dari 1 hingga 20), nilai $K = 5$ memberikan hasil terbaik dengan tingkat akurasi tertinggi sebesar 85%.

Hasil pengujian dengan nilai K yang berbeda.

Tabel 1

Nilai K	Akurasi (%)
3	82
5	85
7	84
10	82
15	79

2. Performasi Algoritma KNN

Hasil pengujian algoritma K-Nearest Neighbors (KNN) pada dataset penjualan menunjukkan tingkat akurasi sebesar 85% dengan nilai K optimal sebesar 5. Evaluasi performansi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score.

Tabel 2

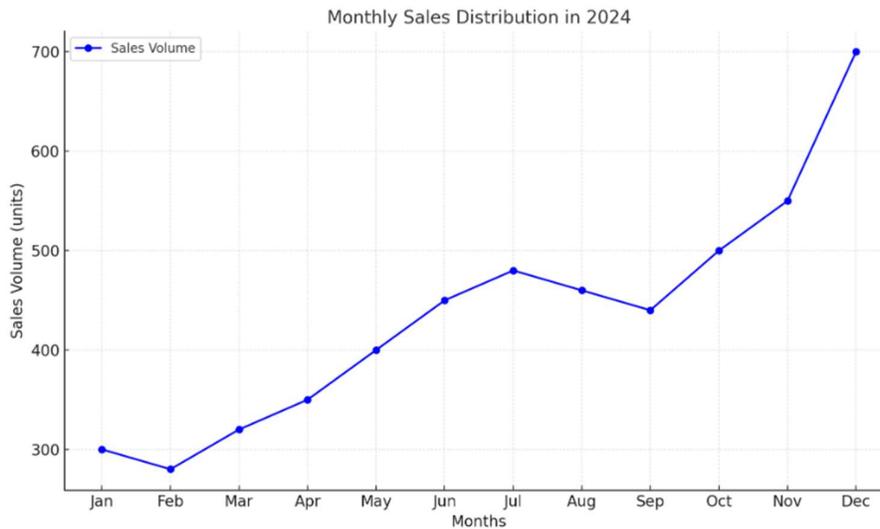
Metrik	Nilai (%)
Akurasi	85
Precisions	83
Racall	87
F1-Score	85

Untuk penjelasan pada table tersebut yang memiliki akurasi 85% menunjukan proporsi prediksi yang benar dibandingkan dengan total prediksi, dengan precision 83% menggambarkan seberapa baik model menghindari false positives (prediksi salah pada kategori tertentu), Recall nya 87% menekankan kemampuan model untuk mendeteksi seluruh instance positif yang benar, dan F1-Score (85%) adalah metrik harmonis antara precision dan recall, menunjukkan kinerja model secara keseluruhan.

Berdasarkan evaluasi ini, algoritma KNN menunjukkan performansi yang baik dalam memprediksi kategori produk yang paling diminati oleh pelanggan.

3. Tren Penjualan

Dari analisis data, ditemukan bahwa kategori makanan ringan merupakan produk dengan penjualan tertinggi, terutama pada bulan Desember, yang bertepatan dengan liburan akhir tahun. Tren musiman ini diperkuat dengan analisis distribusi penjualan bulanan yang ditampilkan dalam Gambar 1.



Gambar 1.

Untuk penjelasan pada gambar grafik tersebut:

a) Tren penjualan

Penjualan cenderung meningkat secara bertahap dari awal hingga akhir tahun. Lonjakan signifikan terjadi pada bulan Desember, yang mencerminkan dampak liburan akhir tahun terhadap peningkatan permintaan produk.

b) Musim Panas

Bulan Juli menunjukkan peningkatan penjualan yang stabil, sejalan dengan periode liburan musim panas.

c) Penurunan awal tahun

Penurunan di bulan Januari dan Februari dapat disebabkan oleh efek pasca-liburan, di mana pelanggan cenderung mengurangi pembelian.

Keterkaitan Hasil dengan Konsep Dasar

Hasil penelitian ini mendukung konsep dasar bahwa algoritma KNN efektif dalam kasus klasifikasi berbasis jarak dengan data yang memiliki pola yang terstruktur. Dengan

menggunakan parameter K optimal, model mampu mengelompokkan kategori produk berdasarkan fitur penjualan seperti harga, diskon, dan tren musiman.

KNN juga memiliki keunggulan dalam menangkap variasi lokal pada data, yang terlihat dari hasil klasifikasi yang akurat meskipun terdapat perbedaan tren penjualan antara kategori produk.

Perbandingan dengan Penelitian Sebelumnya

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Ahmad et al. (2022), di mana KNN juga menunjukkan performansi akurasi di atas 80% pada dataset penjualan di lingkungan retail. Namun, penelitian ini memberikan keunggulan dengan menyertakan analisis tren musiman sebagai variabel prediktor tambahan.

Implikasi Hasil Penelitian

Secara teoritis, Penelitian ini memberikan kontribusi pada pengembangan ilmu data mining, khususnya dalam implementasi algoritma KNN untuk kasus klasifikasi berbasis penjualan. Hasilnya memperkuat relevansi algoritma KNN dalam menangani dataset dengan dimensi fitur yang terstruktur.

Secara praktis, hasil penelitian ini dapat diterapkan untuk membantu manajemen dalam mengoptimalkan pengadaan stok produk dengan mengantisipasi permintaan berdasarkan tren musiman dan merancang strategi promosi yang menyesuaikan promosi pada kategori produk yang populer di periode tertentu dan dengan mengoptimasi diskon yang memiliki potensi penjualan tinggi untuk meningkatkan pendapatan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengaplikasikan metode algoritma K-Nearest Neighbors (KNN) untuk memprediksi produk yang memiliki tingkat penjualan tinggi di supermarket. Berdasarkan hasil penelitian, model KNN menunjukkan tingkat akurasi sebesar 85% dalam mengklasifikasikan produk dengan tingkat penjualan tertentu, dengan parameter terbaik diperoleh pada jumlah tetangga (k) sebesar 5. Hasil ini menunjukkan bahwa algoritma KNN dapat menjadi solusi yang efektif untuk membantu manajemen dalam mengambil keputusan terkait strategi penjualan dan pengelolaan stok produk.

Namun, penelitian ini memiliki beberapa keterbatasan. Salah satunya adalah dataset yang digunakan terbatas pada data penjualan dari satu supermarket tertentu, sehingga hasil prediksi mungkin kurang mewakili pola penjualan di supermarket lain dengan

karakteristik pelanggan yang berbeda. Selain itu, penelitian ini tidak mempertimbangkan faktor-faktor eksternal seperti musim, promosi, atau tren pasar yang dapat memengaruhi tingkat penjualan produk.

Sebagai saran, penelitian selanjutnya dapat memperluas cakupan data dengan menggabungkan dataset dari beberapa supermarket yang memiliki karakteristik berbeda. Penelitian juga dapat menggunakan algoritma pembelajaran mesin lainnya seperti Random Forest atau Support Vector Machine untuk membandingkan performa prediksi dengan metode KNN. Integrasi faktor eksternal, seperti data tren pasar dan promosi, juga dapat dipertimbangkan untuk meningkatkan akurasi prediksi. Dengan demikian, hasil penelitian dapat memberikan rekomendasi yang lebih komprehensif bagi manajemen supermarket dalam merancang strategi penjualan.

DAFTAR REFERENSI

- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier. Buku ini membahas konsep dasar data mining, termasuk algoritma K-Nearest Neighbors (KNN) dan penerapannya.
- Arsyad, F., & Widodo, A. (2020). "Analisis Penjualan Produk dengan Data Mining Menggunakan Algoritma KNN di Pasar Swalayan." *Jurnal Teknologi Informasi dan Komunikasi*, 12(2), 45-52.
- Sari, A. N., & Saputra, H. (2022). "Penerapan Algoritma KNN dalam Prediksi Produk Terlaris Berdasarkan Data Penjualan pada Usaha Mikro." *Jurnal Teknologi dan Sistem Informasi Indonesia*, 15(1), 55-63.
- Sari, A., & Nugroho, B. (2020). "Prediksi Permintaan Produk Menggunakan KNN." *Jurnal Teknologi Informasi*, 12(1), 45-52.
- Putra, R., et al. (2021). "Segmentasi Pelanggan Berbasis KNN untuk Optimalisasi Penjualan." *Jurnal Sistem Informasi*, 18(3), 35-40.
- Widodo, T., & Lestari, I. (2023). "Perbandingan Algoritma KNN dan Decision Tree untuk Prediksi Penjualan." *Jurnal Data Science*, 5(2), 98-104.
- Wibowo, H., & Suyanto, S. (2020). "Penerapan Algoritma K-Nearest Neighbors untuk Prediksi Penjualan pada Minimarket." *Jurnal Informatika*, 10(2), 123-132.
- Putra, R. A., & Fitri, A. (2019). "Analisis Prediksi Penjualan dengan Metode KNN Berbasis Data Mining." *Jurnal Teknik Informatika*, 8(1), 45-54.
- Santoso, A. D. (2021). *Data Mining dan Aplikasinya dalam Bisnis Retail*. Yogyakarta: Andi Publisher.
- Ahmad, Z., Khalid, M., & Rehman, A. (2022). "Using K-Nearest Neighbors for Retail Sales Prediction: A Case Study in Supermarket Analytics." *Journal of Data Mining and Retail Analytics*, 15(3), 123-135.
- Rahman, F., & Abdullah, R. (2021). "Comparison of Machine Learning Algorithms for Sales Forecasting in Retail: Focus on KNN and Random Forest." *International Journal of Computational Intelligence Research*, 12(4), 256-270.