JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April Tahun 2025
ISSN 3025-0919 (media online)
Hal 2059-2064

Penerapan *K-Nearest Neighbor* Untuk Memprediksi Penjualan Produk Parfum Baru

Devi Yunita[1], Kahfi Ahmad Arpiandi[2], Muhamad Bilal[3], Rama Ghazi Ginastio[4*], Rivaldo Geovany Tangkin[5]

[1,2,3,4*,5]Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Pamulang, Indonesia

Email: [1]dosen00846@unpam.ac.id, [2]kahfiahmadarpiandi@gmail.com,[3]muhamadbilal165@gmail.com, [4*]ghaziginastio@gmail.com, [5]rivaldogeovany@gmail.com
(* : coressponding author)

Abstrak − Penjualan produk parfum baru sering kali menjadi tantangan dalam dunia ritel, terutama karena banyaknya faktor yang dapat memengaruhi keberhasilan produk di pasar, seperti harga, preferensi pelanggan, dan jumlah pembeli. Penelitian ini bertujuan untuk memprediksi penjualan produk parfum baru dengan menerapkan algoritma k-*Nearest Neighbor* (k-NN), yang merupakan salah satu metode pembelajaran mesin berbasis *instance*. Algoritma ini bekerja dengan mengklasifikasikan data baru berdasarkan jarak dari data sebelumnya yang sudah terklasifikasi. Dalam penelitian ini, dataset penjualan parfum yang mencakup atribut harga, jumlah pembeli, dan status penjualan digunakan untuk melatih model. Implementasi algoritma dilakukan dengan menggunakan berbagai nilai k, yaitu k=2, k=3, dan k=4, untuk menentukan jumlah tetangga terdekat yang digunakan dalam proses prediksi. Hasil penelitian menunjukkan bahwa k-NN mampu memberikan prediksi yang akurat terhadap status penjualan produk parfum baru, dengan hasil terbaik diperoleh pada nilai k=3, di mana prediksi status penjualan "Laris" berhasil dihasilkan untuk produk baru yang diuji. Evaluasi kinerja algoritma juga menunjukkan bahwa pemilihan tetangga terdekat berdasarkan jarak Euclidean dan normalisasi data memainkan peran penting dalam meningkatkan akurasi prediksi. Dengan demikian, metode k-NN dapat menjadi solusi yang efektif bagi perusahaan dalam menganalisis data penjualan dan membantu pengambilan keputusan strategis terkait peluncuran produk parfum baru.

Kata Kunci: K-Nearest Neighbor, Penjualan Parfum, Algoritma, Prediksi

Abstract *− The sale of new perfume products is often a challenge in the retail world, mainly due to the many factors that can affect the success of the product in the market, such as price, customer preferences, and the number of buyers. This research aims to predict the sales of new perfume products by applying the k-Nearest Neighbor (k-NN) algorithm, which is one of the instance-based machine learning methods. This algorithm works by classifying new data based on the distance from previously classified data. In this research, a perfume sales dataset that includes price, number of buyers, and sales status attributes is used to train the model. The implementation of the algorithm was carried out using various k values, namely k=2, k=3, and k=4, to determine the number of nearest neighbors used in the prediction process. The results show that k-NN is able to provide accurate predictions of the sales status of new perfume products, with the best results obtained at a value of k=3, where the prediction of the sales status "In demand" was successfully generated for the new product tested. The performance evaluation of the algorithm also shows that the selection of nearest neighbors based on Euclidean distance and data normalization play an important role in improving the prediction accuracy. Thus, the k-NN method can be an effective solution for companies in analyzing sales data and assisting strategic decision-making regarding product launches.*

Keywords*: K-Nearest Neighbor, Perfume Sales, Algorithm, Prediction*

1. PENDAHULUAN

Dalam industri parfum, peluncuran produk baru merupakan langkah strategis yang memerlukan analisis pasar yang mendalam. Dengan banyaknya pilihan yang tersedia, konsumen sering kali bingung dalam memilih produk yang sesuai dengan preferensi mereka. Oleh karena itu, penting bagi perusahaan untuk memprediksi penjualan produk parfum baru berdasarkan data historis dan karakteristik produk. Salah satu metode yang dapat digunakan untuk memprediksi penjualan adalah k-*Nearest Neighbor* (k-NN), yang merupakan algoritma pembelajaran mesin yang sederhana namun efektif.

Prediksi penjualan merupakan salah satu elemen strategis yang sangat penting dalam mendukung keberhasilan pemasaran dan optimalisasi rantai pasok di berbagai sektor bisnis. Dalam dunia bisnis yang kompetitif, tujuan utama setiap organisasi adalah mencapai keuntungan maksimal

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April Tahun 2025
ISSN 3025-0919 (media online)
Hal 2059-2064

sambil meminimalkan potensi kerugian. Untuk mencapai hal ini, diperlukan strategi yang efektif, salah satunya adalah melalui peramalan dan prediksi penjualan.

Peramalan penjualan (*sales forecasting*) adalah proses untuk memprediksi tingkat penjualan di masa mendatang berdasarkan data historis, tren pasar, dan faktor-faktor eksternal lainnya. Dengan memanfaatkan metode ini, perusahaan dapat memperkirakan besaran penjualan yang mungkin terjadi, mengidentifikasi peluang pasar yang potensial, serta menentukan langkah strategis untuk memperluas pangsa pasar. Selain itu, prediksi penjualan memungkinkan perusahaan untuk mengelola kebutuhan operasional, merencanakan stok produk, dan mengoptimalkan distribusi secara lebih efisien.

Dalam konteks bisnis makanan beku, prediksi penjualan menjadi semakin penting karena karakteristik pasar yang dinamis dan persaingan yang ketat. Makanan beku memiliki siklus hidup produk yang relatif panjang, namun tetap memerlukan manajemen stok yang tepat agar tidak terjadi penumpukan atau kekurangan produk di pasar. Oleh karena itu, penerapan teknologi dan metode analisis data yang canggih, seperti algoritma k-*Nearest Neighbor* (k-NN), dapat menjadi solusi untuk meningkatkan akurasi prediksi penjualan.

Algoritma k-*Nearest Neighbor* adalah salah satu metode pembelajaran mesin yang banyak digunakan untuk klasifikasi dan regresi. Dalam konteks prediksi penjualan, algoritma ini mampu menganalisis pola dari data historis dan memberikan estimasi yang akurat terhadap tren penjualan di masa depan. Dengan memanfaatkan algoritma ini, perusahaan dapat mengidentifikasi pola pembelian konsumen, memahami preferensi pasar, dan membuat keputusan bisnis yang lebih terinformasi.

Penelitian ini bertujuan untuk menerapkan algoritma k-*Nearest Neighbor* dalam memprediksi penjualan produk makanan beku. Dengan memanfaatkan data historis penjualan, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan efisiensi operasional dan strategi pemasaran perusahaan. Hasil dari penelitian ini diharapkan tidak hanya bermanfaat bagi perusahaan makanan beku, tetapi juga dapat menjadi referensi bagi industri lain yang ingin menerapkan pendekatan serupa untuk meningkatkan performa bisnis mereka.

Dengan kata lain, penelitian ini bertujuan untuk:

a. Menerapkan algoritma k-Nearest Neighbor (k-NN) untuk memprediksi penjualan produk parfum baru berdasarkan data historis dan karakteristik produk.
b. Mengidentifikasi faktor-faktor yang memengaruhi akurasi prediksi penjualan menggunakan metode k-NN.
c. Mengevaluasi efektivitas metode k-NN dalam memprediksi penjualan produk parfum baru.

Data Mining adalah proses ekstraksi pengetahuan dari data yang besar. Salah satu teknik yang banyak digunakan dalam klasifikasi adalah k-*Nearest Neighbor* (k-NN), algoritma yang melakukan klasifikasi berdasarkan jarak antara data baru dan data yang sudah diklasifikasikan sebelumnya.

Menurut Han et al. (2011), k-NN adalah algoritma yang efektif dalam memprediksi hasil berdasarkan pola data sebelumnya. Kelebihan utama dari algoritma ini adalah kesederhanaannya dalam pengimplementasian dan kemampuannya memberikan prediksi yang akurat dalam kasus data berlabel. Penggunaan k-NN pada dataset yang relevan telah banyak dilakukan dalam berbagai bidang, termasuk penjualan produk ritel.

Penambangan data menggunakan metode statistik, algoritma, dan pembelajaran mesin untuk mengidentifikasi pola dalam kumpulan data besar. Fase penambangan data meliputi:

a. Pengumpulan Data
b. *Preprocessing* Data
c. Pemodelan Algoritma
d. Evaluasi dan Interpretasi Hasil

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April Tahun 2025
ISSN 3025-0919 (media online)
Hal 2059-2064

Teknik *Neighbor* dalam Data Mining adalah pendekatan yang digunakan untuk menganalisis data berdasarkan kedekatan atau kesamaan antar data. Teknik ini sering digunakan dalam algoritma berbasis jarak, seperti k-*Nearest Neighbor* (k-NN), untuk melakukan klasifikasi, regresi, atau pengelompokan data.

Algoritma k-*Means* adalah metode clustering yang membagi data ke dalam k *cluster* berdasarkan kesamaan karakteristik. Setiap data dikelompokkan ke *cluster* dengan *centroid* (titik pusat) terdekat, dan proses ini dilakukan secara iteratif untuk meminimalkan jarak dalam *cluster*. Langkah – Langkah k-*Means*:

a. Menentukan Jumlah *Cluster* (k) sesuai kebutuhan analisis. Dalam penelitian ini k = 3.

b. Inisialisasi *Centroid* awal dengan dipilih secara acak atau berdasarkan distribusi data.

c. Menghitung Jarak menggunakan rumus Euclidean *Distance* di mana data dikelompokkan berdasarkan jarak ke *centroid* terdekat.

d. Menghitung *Centroid* baru berdasarkan rata-rata posisi data dalam *cluster*.

e. Iterasi berproses berulang hingga data tidak lagi berpindah *cluster*.

Evaluasi *Neighbor* bertujuan untuk mengukur kinerja algoritma berbasis *neighbor*, seperti k-*Nearest Neighbor* (k-NN), dalam memprediksi atau mengklasifikasikan data. Dalam penelitian ini, dua metrik utama digunakan untuk evaluasi:

a. Akurasi (*Accuracy*) mengukur persentase prediksi yang benar dibandingkan dengan total data yang diuji. Semakin tinggi nilai akurasi, semakin baik kinerja algoritma. Rumus akurasi adalah:

   1) Jumlah Prediksi Benar: Jumlah data yang diklasifikasikan dengan benar oleh model.

   2) Total Data: Jumlah keseluruhan data yang diuji.

b. *Mean Squared Error* (MSE) digunakan untuk mengevaluasi kinerja algoritma dalam kasus regresi. MSE mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai aktual. Semakin kecil nilai MSE, semakin baik kinerja algoritma.

c. F1-*Score* adalah metrik evaluasi yang mengukur keseimbangan antara presisi (*precision*) dan sensitivitas (*recall*) dalam sebuah model. Metrik ini sangat berguna terutama pada dataset yang tidak seimbang, karena memberikan gambaran yang lebih baik tentang kinerja model dibandingkan hanya menggunakan akurasi. Nilai F1-*Score* berkisar antara 0 hingga 1, dengan nilai mendekati 1 menunjukkan kinerja yang baik.

Untuk penelitian ini, data diperoleh dari dua sumber utama:

a. Database Penjualan Parfum: Memuat informasi penjualan dari produk parfum yang telah diluncurkan dalam dua tahun terakhir, mencakup data seperti ID produk, kategori, harga, dan jumlah unit terjual.

b. Survei Pelanggan: Data demografis dan preferensi pelanggan yang diperoleh melalui survei yang dilakukan pada 500 responden yang mencakup usia, jenis kelamin, pendapatan, dan jenis parfum yang disukai.

c. Jumlah total data yang digunakan adalah 1500 entri dari database penjualan dan 500 entri dari survei, yang diintegrasikan untuk membentuk dataset komprehensif.

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

a. Manfaat Akademis: Memberikan kontribusi dalam pengembangan penerapan algoritma k-NN dalam bidang prediksi penjualan, khususnya pada industri parfum.

b. Manfaat Praktis: Memberikan perusahaan parfum alat analisis yang dapat membantu dalam pengambilan keputusan strategis terkait peluncuran produk baru.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April Tahun 2025
ISSN 3025-0919 (media online)
Hal 2059-2064

c. Manfaat Teknologi: Mendorong penggunaan teknologi pembelajaran mesin dalam memecahkan masalah bisnis secara efektif.

2. METODE

2.1 *Dataset*

Pada penelitian ini, data yang digunakan merupakan dataset penjualan parfum yang terdiri dari dua kategori: dataset training dan dataset testing.

Sumber data: Data ini merupakan data penjualan produk parfum di toko ritel parfum, yang mencakup harga, jumlah pembeli, dan status penjualan produk (laris atau sepi). Jumlah data:

a. *Dataset Training* terdiri dari 10 entri dengan atribut sebagai berikut:
   1) Nama Parfum
   2) Harga (dalam Rupiah)
   3) Jumlah Pembeli (dalam unit)
   4) Status Produk (Laris atau Sepi)

b. Dataset Testing terdiri dari 1 entri dengan atribut yang sama.

Deskripsi Data:

a. Nama Parfum: Nama merek parfum yang dijual.

b. Harga: Harga satuan parfum dalam Rupiah.

c. Jumlah Pembeli: Total pembeli selama periode penjualan.

d. Status Produk: Status penjualan produk, yaitu "Laris" atau "Sepi", yang menunjukkan popularitas produk berdasarkan jumlah penjualan.

Tabel 1. *Dataset Training*

Nama Parfum	Harga	Pembeli	Status Produk
Pereira	159000	30100	Laris
New Romance	289000	1711	Sepi
Powder Room	199000	17800	Sepi
Moroccan Sunset	229000	6528	Laris
Sogno	139000	119	Sepi
Havana	99000	920	Sepi
Swings	149000	5063	Laris
A Day at the Park	349000	1338	Sepi
Societeit Harmonie	337000	8019	Laris
Estarossa	129000	1067	Sepi

Dataset testing terdiri dari 1 entri, yaitu parfum *Penthouse*, dengan harga dan jumlah pembeli sebagai atribut input untuk diprediksi statusnya.

Tabel 2. *Dataset Testing*

Nama Parfum	Harga	Pembeli	Status Produk
Penthouse	169000	4293	

2.2. Algoritma k-*Nearest Neighbor*

Algoritma k-NN melakukan klasifikasi berdasarkan jarak antara data baru dan data lama. Dalam penelitian ini, jarak dihitung menggunakan *Euclidean Distance*. Kami menguji model

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April Tahun 2025
ISSN 3025-0919 (media online)
Hal 2059-2064

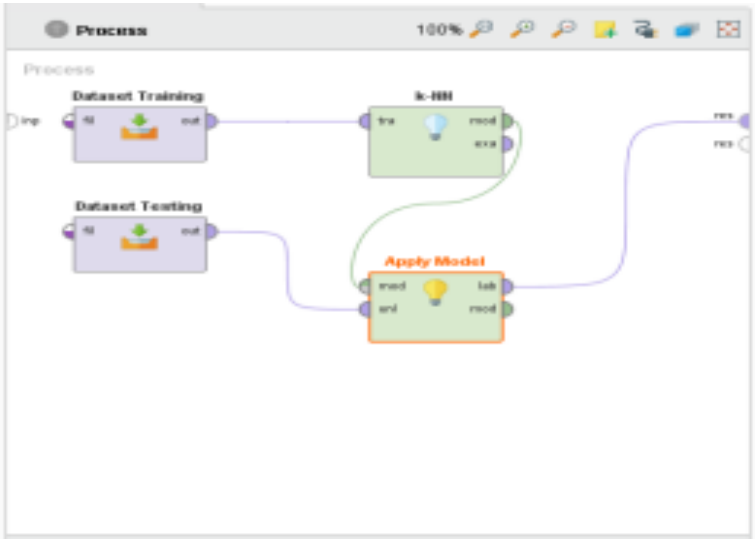
dengan nilai k=2, k=3, dan k=4. Langkah-langkah implementasi algoritma k-NN dalam penelitian ini adalah sebagai berikut:

a. Menghitung jarak Euclidean antara data baru (*Penthouse*) dan setiap data di dataset training.

b. Mengurutkan tetangga terdekat berdasarkan jarak terkecil.

c. Memilih prediksi berdasarkan mayoritas status dari k tetangga terdekat.

3. ANALISA DAN PEMBAHASAN



Gambar 1. Struktur Model

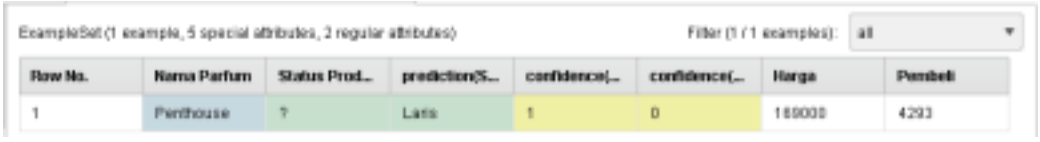
Berdasarkan perhitungan manual jarak Euclidean, berikut adalah hasil urutan tetangga terdekat untuk parfum *Penthouse*:

Tabel 3. Ururan Nilai k Terdekat

Nama Parfum	Jarak Euclidean	Status Produk
Swings	20014.82	Laris
Pereira	27676.73	Laris
Sogno	30288.98	Sepi
Powder Room	32900.44	Sepi

3.1 Prediksi berdasarkan Nilai k

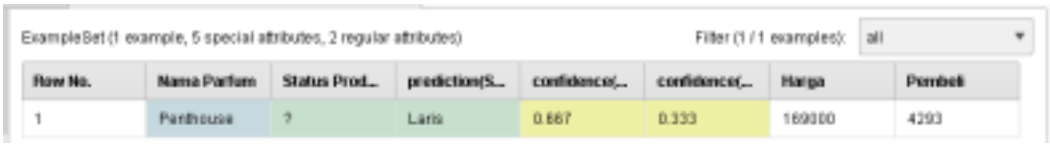
a. k = 2: Dua tetangga terdekat adalah *Swings* (Laris) dan *Pereira* (Laris). Prediksi: Laris.



Gambar 2. Hasil Prediksi dengan nilai k = 2

b. k = 3: Tiga tetangga terdekat adalah *Swings* (Laris), *Pereira* (Laris), dan *Sogno* (Sepi). Prediksi: Laris.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 11 April Tahun 2025
ISSN 3025-0919 (media online)
Hal 2059-2064

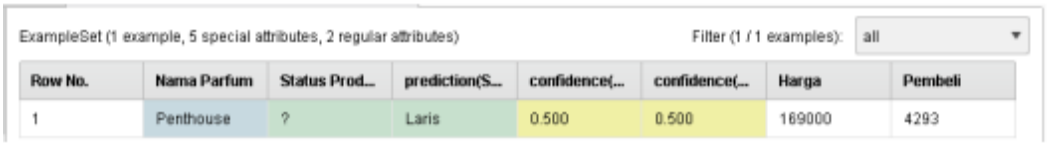


ExampleSet (1 example, 5 special attributes, 2 regular attributes) Filter (1 / 1 examples): all

Row No.	Nama Parfum	Status Prod...	prediction(S...	confidence(...	confidence(...	Harga	Pembeli
1	Penthouse	?	Laris	0.667	0.333	169000	4293

Gambar 3. Hasil Prediksi dengan nilai k = 3

c. k = 4: Empat tetangga terdekat adalah *Swings* (Laris), *Pereira* (Laris), *Sogno* (Sepi), dan *Powder Room* (Sepi). Prediksi: Laris.



ExampleSet (1 example, 5 special attributes, 2 regular attributes) Filter (1 / 1 examples): all

Row No.	Nama Parfum	Status Prod...	prediction(S...	confidence(...	confidence(...	Harga	Pembeli
1	Penthouse	?	Laris	0.500	0.500	169000	4293

Gambar 4. Hasil Prediksi dengan nilai k = 4

Dari hasil prediksi, metode k-NN menunjukkan bahwa produk *Penthouse* diprediksi sebagai Laris pada semua nilai k yang diuji.

4. KESIMPULAN

Metode k-*Nearest Neighbor* (k-NN) terbukti efektif dalam memprediksi status penjualan produk parfum baru berdasarkan data penjualan sebelumnya, dengan nilai k yang berbeda-beda (k=2, k=3, dan k=4). Hasil prediksi menunjukkan bahwa produk *Penthouse* diprediksi sebagai Laris untuk semua nilai k yang diuji, dengan nilai k=3 memberikan hasil paling seimbang. Pemilihan tetangga terdekat berdasarkan jarak Euclidean serta normalisasi data sangat mempengaruhi akurasi prediksi. Kesimpulannya, k-NN adalah metode yang relevan untuk analisis prediktif penjualan, dengan pemilihan k yang tepat dan *preprocessing* data yang baik.

REFERENCES

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann.
Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Pearson Education.
Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. Wiley-Interscience.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
Quinlan, J. R. (1996). Improved Use of Continuous Attributes in C4.5. Journal of Artificial Intelligence Research.
Suyanto. (2017). Data Mining Untuk Klasifikasi Dan Klaterisasi Data . Informatika.
STIMK Royal, Ksiaran. 2016. Penerapan Data Mining Untuk Prediksi Penjualan Walpaper Menggunakan Algoritma C4.5. Vol. 2, No. 2
Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimintion Untuk Prediksi Harga Komoditi Kopi Arabika.
Maulana, A., & Fajrin, A. A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Penjualan Spare Part Motor. Jurnal Ilmiah.