

# Implementasi Data Mining Dalam Memprediksi Penjualan Parfum Terlaris Menggunakan Metode K-Nearest Neighbor

Mhd. Angga Sabda\*, Suhardi

Fakultas Ilmu Komputer, Program Studi Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: <sup>1,\*</sup>aanggasabda@gmail.com, <sup>2</sup>suhardi@uinsu.ac.id

Email Penulis Korespondensi: aanggasabda@gmail.com

Submitted: 19/12/2023; Accepted: 29/12/2023; Published: 29/12/2023

**Abstrak**—Ahlinyaparfum adalah toko parfum yang menyediakan berbagai macam minyak wangi. Untuk menjalankan bisnisnya, Ahlinyaparfum harus mengirimkan varian parfum ke tokonya dari tempat pembuatan parfum tersebut, yang memerlukan biaya pengiriman. Seringkali, varian parfum yang ditawarkan oleh Ahlinyaparfum tidak sesuai dengan keinginan pelanggan, yang berdampak negatif pada jumlah penjualan toko. Proses prediksi penjualan terlaris diperlukan berdasarkan data penjualan sebelumnya untuk membantu toko mengetahui parfum mana yang paling diminati pelanggan dan tingkat penjualan terlaris di masa depan. Dengan menerapkan data mining menggunakan metode K-Nearest Neighbor, penelitian ini bertujuan untuk mengatasi masalah tersebut. Metode ini diuji menggunakan data penjualan parfum bulan Januari hingga Juni 2023 dengan total 215 data. Untuk menguji dan memastikan kinerjanya dengan bantuan aplikasi Jupyter Notebook dengan Python. Proses prediksi penjualan parfum pada bulan berikutnya menggunakan parameter  $k = 3$  dengan perhitungan euclidean distance. Hasil prediksi terlaris bulan ke 7 adalah Aigner Blue Emotion dengan total penjualan 153 ml. Berdasarkan evaluasi algoritma, nilai rata-rata keseluruhan MSE adalah 0,52, yang menunjukkan bahwa hasilnya sangat baik dalam menentukan penjualan parfum bulan depan. Ini disebabkan oleh fakta bahwa hasil perhitungan semakin dekat dengan nilai aktual semakin rendah nilai MSE.

**Kata Kunci:** Prediksi Penjualan; Parfum; Data Mining; K-Nearest Neighbor; MSE

**Abstract**—Ahlinyaparfum is a perfume shop that provides various kinds of fragrance oils. To run his business, Ahlinyaparfum must send perfume variants to his shop from the place where the perfume is made, which requires shipping costs. Often, the perfume variants offered by Ahlinyaparfum do not match the wishes of customers, which has a negative impact on the number of store sales. The best-selling prediction process is needed based on previous sales data to help stores know which perfumes are most popular with customers and the level of best-selling in the future. By applying data mining using the K-Nearest Neighbor method, this research aims to overcome this problem. This method was tested using perfume sales data from January to June 2023 with a total of 215 data. To test and ensure its performance with the help of the Jupyter Notebook application with Python. The process of predicting perfume sales for the next month uses the parameter  $k = 3$  with Euclidean distance calculations. The best-selling result predicted for the 7th month is Aigner Blue Emotion with total sales of 153 ml. Based on the evaluation algorithm, the overall average value of MSE is 0.52, which shows that the results are very good in determining next month's perfume sales. This is due to the fact that the calculation results are closer to the actual value the lower the MSE value.

**Keywords:** Sales Predictions; Perfume; Data Mining; K-Nearest Neighbor; MSE

## 1. PENDAHULUAN

Dalam bisnis, hal yang paling diinginkan adalah menghasilkan keuntungan yang paling besar dan kerugian yang paling kecil. Peramalan dan prediksi penjualan adalah salah satu strategi penjualan yang dapat digunakan untuk meningkatkan omzet penjualan. Peramalan (forecasting) penjualan adalah upaya untuk memprediksi penjualan di masa mendatang dan menentukan besaran penjualan dan luas pasar yang dikuasai di masa mendatang. Strategi penjualan ini bertujuan untuk mengetahui penjualan di masa depan dan dapat mempersiapkan kebutuhan dan operasional pelaksanaan bisnis dari hasil prediksi penjualan dengan maksimal [1].

Karena perusahaan memiliki banyak data yang dapat diubah menjadi informasi dan pengetahuan yang bermanfaat, pengetahuan masa depan sangat penting untuk mendukung proses bisnis mereka [2]. Untuk memenuhi keinginan dan kepuasan pelanggan dalam penjualan berikutnya, informasi data sebelumnya sangat berguna [3].

Ahlinyaparfum adalah toko parfum yang menyediakan berbagai macam minyak wangi. Toko ini memiliki banyak pilihan parfum dan kemasan. Ketika dijual, pelanggan dapat membeli parfum bersama botolnya dan kemudian hanya membeli isi ulangnya. Untuk menjalankan bisnisnya, Ahlinyaparfum harus mengirimkan varian parfum ke tokonya dari tempat pembuatan parfum tersebut, yang memerlukan biaya pengiriman. Seringkali, varian parfum yang ditawarkan oleh Ahlinyaparfum tidak sesuai dengan keinginan pelanggan, yang berdampak negatif pada jumlah penjualan toko. Ini karena belum ada sistem yang dapat mengetahui varian parfum mana yang paling diminati pelanggan dan belum ada cara yang akurat untuk mengetahui parfum mana yang paling laris di toko tersebut. Ahlinyaparfum harus menggunakan kemajuan teknologi saat ini untuk mengatasi masalah parfum mana yang paling diminati pelanggan dan menentukan tingkat penjualan terlaris di masa depan. Oleh karena itu, untuk mengetahui berapa banyak penjualan di masa depan dan jenis parfum mana yang paling diminati oleh pelanggan, proses prediksi penjualan parfum terlaris harus dilakukan dalam membantu Ahlinyaparfum mempersiapkan kebutuhan dan menjalankan bisnisnya dengan baik [4].

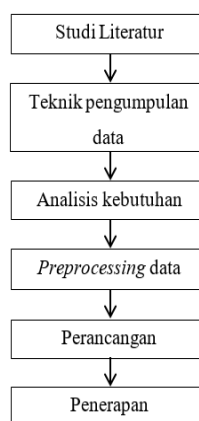
Dalam penelitian sebelumnya, telah dilakukan prediksi penjualan produk makanan kering terlaris menggunakan data penjualan tahun 2021-2022 menggunakan atribut 5 bulan penjualan dan target penjualan bulan

6. Dalam prosesnya menggunakan metode K-Nearest Neighbor (K-NN), dengan perhitungan jarak menggunakan euclidean, dan digunakan  $k = 3$ , hasil yang didapat menunjukkan bahwa, melalui perhitungan K-NN, memiliki nilai Root Mean Square Error (RMSE) terkecil, yaitu 2,869, yang menunjukkan bahwa hasil prediksinya hampir sama dengan hasil sebenarnya, sedangkan peneliti melakukan prediksi penjualan parfum terlaris dengan data penjualan parfum bulan 1-6 tahun 2023 menggunakan atribut penjualan 6 bulan dan target dari penjualan rata-rata 6 bulan [1]. Pada penelitian lainnya, tentang penggunaan data mining menggunakan metode K-NN untuk memprediksi harga minyak mentah dunia berdasarkan data harga minyak tahun 2018-2022 dengan membandingkan hasil model K-NN tunggal dengan hasil model ensemble K-NN, hasil akhir keduanya mendapatkan akurasi baik dalam memprediksi harga minyak, sedangkan peneliti akan melakukan model K-NN tunggal menggunakan nilai  $k = 3$  [5]. Selanjutnya pada penelitian lain, melakukan prediksi penjualan sparepart mobil terlaris berdasarkan penjualan 2019-2020 dengan algoritma K-NN menggunakan software RapidMiner, hasil yang didapat memperoleh akurasi 80%, sedangkan peneliti menggunakan Python untuk memprediksi penjualan parfum terlaris di masa bulan berikutnya [3]. Berdasarkan penelitian sebelumnya metode K-Nearest Neighbor sangat efektif dalam memprediksi data besar [6], jadi data mining menggunakan metode K-NN digunakan dalam penelitian ini untuk membuat prediksi penjualan parfum terlaris.

Penelitian ini menggunakan aplikasi Jupyter Notebook dan bahasa pemrograman Python untuk melakukan proses perhitungan K-Nearest Neighbor. Jupyter Notebook adalah software pemrograman interaktif untuk menggabungkan kode-kode, teks naratif, gambar grafis, dan visualisasi [7]. Jupyter Notebook adalah aplikasi text editor berbasis web yang berjalan di localhost [8]. Python adalah bahasa pemrograman tingkat tinggi yang sering digunakan oleh perusahaan dan pengembang besar untuk membuat aplikasi web, desktop, dan seluler. Karena mudah digunakan dan memiliki perpustakaan yang besar, Python menjadi bahasa pemrograman tingkat tinggi yang banyak digunakan di bidang pendidikan dan industri [9].

## 2. METODOLOGI PENELITIAN

Peneliti akan melakukan beberapa langkah dalam penelitian ini untuk memprediksi penjualan parfum terlaris dengan K-Nearest Neighbor yang dapat dilihat pada Gambar 1.



**Gambar 1.** Kerangka Penelitian

### 2.1. Studi Literatur

Kegiatan penelitian literatur dilakukan untuk mengumpulkan informasi tentang topik penelitian peneliti dari berbagai literatur, seperti buku cetak, e-book, jurnal, dan karya ilmiah lainnya.

### 2.2. Teknik Pengumpulan Data

Dalam metode pengumpulan data, ada dua tahap. Pertama, pemilik usaha Ahlinyaparfum diwawancarai untuk mendapatkan informasi tentang barang yang dijual, seperti parfum, riwayat penjualan, dan informasi lain yang terkait dengan penelitian. Untuk mengetahui masalah sebenarnya, observasi dilakukan bersamaan dengan wawancara untuk mendukung hasil wawancara.

### 2.3. Analisis Kebutuhan

Untuk melakukan penelitian ini dengan baik, beberapa analisis diperlukan. Data penjualan parfum adalah fokus utama penelitian ini, meliputi penjualan bulan Januari hingga bulan Juni 2023, nama parfum dan kuantiti parfum terjual tiap bulan.

Selain itu, penelitian ini menggunakan algoritma K-NN untuk melakukan proses dari data yang sudah ada yaitu penjualan parfum bulan Januari hingga bulan Juni 2023 dan membuat prediksi penjualan terlaris bulan berikutnya menggunakan bahasa pemrograman Python.

## 2.4. Preprocessing Data

Tahap preprocessing data dilakukan setelah pengumpulan dan analisis data yang diperlukan. Untuk mendapatkan model yang sangat akurat, proses preprocessing data sangat penting. [10]. Dataset dari berbagai sumber sering mengandung kesalahan, nilai yang kosong, dan ketidakkonsistenan, sehingga diperlukan preprocessing untuk mengubah data tersebut menjadi data yang dapat diolah dengan data mining [11].

Proses preprocessing data dilakukan dalam tiga tahap yakni seleksi data memilih beberapa atribut saja seperti nama parfum dan kuantiti parfum terjual tiap bulannya yang kemudian dilakukan pembersihan data duplikasi, data yang tidak perlu, dan missing value atau nilai yang hilang pada data. Hasil data cleaning yang dilakukan dari 485 data menjadi sebanyak 215 data. Setelah itu dilakukan penyesuaian data ke dalam bentuk yang terpisah, yaitu pembagian data training dan data testing.

## 2.5. Perancangan

Penelitian ini, atribut yang dipilih seperti nama parfum, jumlah parfum yang terjual, dan target penjualan bulan ke-7 dihitung berdasarkan rata-rata penjualan bulan 1 hingga bulan 6. Total dataset yang digunakan sebanyak 215 data, kemudian data tersebut dibagi dalam penelitian ini menggunakan rasio 80:20. Selanjutnya menghitung jarak dari tetangga terdekat menggunakan Euclidean Distance dan menentukan nilai k untuk mengukur jarak terkecil berdasarkan mayoritas nilai. Flowchart adalah alur kerja menggunakan simbol-simbol yang dibuat untuk menggambarkan langkah-langkah dari program komputer [12]. Gambar 2 menunjukkan flowchart yang telah dirancang.



**Gambar 2.** Flowchart K-Nearest Neighbor

Dalam Gambar 2, pertama-tama dimasukkan data penjualan parfum, kemudian ditampilkan. Selanjutnya, untuk menghitung jarak antara data baru dan jumlah tetangga terdekat, diurutkan dari nilai terkecil ke nilai terbesar, rumus jarak euclidean distance digunakan. Selanjutnya, tetapkan hasil prediksi penjualan parfum berdasarkan nilai k, dan selesai.

## 2.6. Penerapan Algoritma K-Nearest Neighbor

Metode klasifikasi dan regresi objek baru yang didasarkan pada tetangga terdekat (K) adalah algoritma K-Nearest Neighbor [13]. Perhitungan K-NN menggunakan rumus euclidean distance, yang merupakan rumus yang paling sederhana yang sering digunakan dalam pengimplementasian pencarian jarak. [14]. Nilai k menunjukkan jumlah tetangga terdekat yang terlibat dalam membuat prediksi label kelas pada data training [15]. Implementasi algoritma K-Nearest Neighbor yang diterapkan pada data penjualan parfum untuk mengetahui penjualan parfum terlaris bulan selanjutnya berdasarkan variabel yang telah ditentukan yaitu, nama parfum dan kuantiti parfum terjual. Penerapan algoritma ini diharapkan dapat mempermudah dalam mengetahui penjualan parfum terlaris di bulan selanjutnya, sehingga nilai prediksi ini dapat digunakan sebagai panduan untuk target penjualan parfum di bulan berikutnya.

# 3. HASIL DAN PEMBAHASAN

## 3.1. Implementasi Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor memerlukan proses preprocessing data dalam tiga tahap. Pertama, data penjualan parfum dari Januari hingga Juni 2023 dari Ahlinyaparfum diseleksi menjadi nama parfum, kuantiti parfum yang

terjual, dan rata-rata penjualan sebagai data target penjualan bulan ke-7. Tabel 1 menunjukkan data penjualan parfum yang telah diseleksi.

**Tabel 1.** Dataset Penjualan Parfum

No.	Nama Parfum	Bulan 1	Bulan 2	::	Bulan 6	Target (Juli)
1.	Aigner Black Man	71	245	::	116	168,33
2.	Vanilla Choco	0	0	::	0	0,17
3.	Aigner Debut	41	37	::	25	45,83
4.	Aigner In Leather M	0	0	::	5	0,83
5.	Aigner Starlight	0	4	::	0	3,17
6.	Aigner Too Feminim	55	0	::	21	21,00
7.	Vanilla Susu	234	171	::	257	286,17
8.	Alexandria	30	0	::	0	6,33
9.	Alfad	31	30	::	28	49,33
::	::	::	::	::	::	::
211.	Dior Sauvage	22	130	::	128	88,83
212.	Nagita Slavina	57	21	::	111	47,83
213.	Invictus	0	0	::	0	4,17
214.	Sporty Lady	26	54	::	20	28,83
215.	Madina Al-Haramain	0	0	::	3	2,50

Tahapan berikutnya adalah pembersihan data duplikasi, data yang tidak perlu, dan missing value atau nilai yang hilang pada data. Hasil data cleaning yang dilakukan dari 485 data menjadi sebanyak 215 data. Setelah itu data disesuaikan ke dalam bentuk yang berbeda yaitu pembagian data training dan data testing. Data training untuk penjualan parfum dalam penelitian ini dibagi dengan rasio 80:20, yang menghasilkan 172 dari rasio 80%. Data training ini disajikan di Tabel 2.

**Tabel 2.** Data Latih

No.	Nama Parfum	Bulan 1	Bulan 2	::	Bulan 6	Target (Juli)
1.	Aigner Black Man	71	245	::	116	168,33
2.	Vanilla Choco	0	0	::	0	0,17
3.	Aigner Debut	41	37	::	25	45,83
4.	Aigner In Leather M	0	0	::	5	0,83
5.	Aigner Starlight	0	4	::	0	3,17
::	::	::	::	::	::	::
172.	Vanilla Coklat	20	9	::	0	12,50

Sedangkan hasil data uji sebanyak 43 dari rasio 20%. Berikut data uji penjualan parfum pada Tabel 3.

**Tabel 3.** Data Uji

No.	Nama Parfum	Bulan 1	Bulan 2	::	Bulan 6	Target (Juli)
173.	Bodyshop Madagaskar	0	21	::	13	10,67
174.	Khaff	0	0	::	32	5,33
175.	Jaguar Vision	23	56	::	32	63,5
176.	Ariana Grande	25	18	::	32	35,83
177.	Guess Pink	11	5	::	42	28
::	::	::	::	::	::	::
215.	Madina Al-Haramain	0	0	::	3	2,50

Setelah dataset melewati tahap preprocessing data, data sudah siap untuk diolah dengan algoritma K-Nearest Neighbor. Penerapan algoritma K-NN memerlukan tindakan berikut.

1. Menentukan Nilai k

Karena kemungkinan nilai k akan menghasilkan nilai kembar atau tidak berbeda jika nilai k menggunakan nilai genap, maka nilai k harus dipilih menjadi nilai ganjil. Penelitian ini menggunakan nilai k = 3.

2. Perhitungan Jarak

Untuk menghitung jarak antara data training dan data testing, penelitian ini memilih rumus jarak Euclidean. Selanjutnya, perhitungan jarak Euclidean dilakukan dengan menggunakan rumus 1.

$$d_i = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \tag{1}$$

dimana  $x_i$  adalah data latih,  $y_i$  adalah data uji,  $d$  adalah jarak, dan  $k$  adalah jarak [16]. Perhitungan antara jarak data training dan data testing dilakukan dengan menggunakan rumus 1.

$$d_1 = \sqrt{((71 - 0)^2 + (245 - 21)^2 + \dots + (116 - 13)^2)} = 408,137$$

$$d_2 = \sqrt{((0 - 0)^2 + (0 - 21)^2 + \dots + (0 - 13)^2)} = 33,332$$

$$d_3 = \sqrt{((41 - 0)^2 + (37 - 21)^2 + \dots + (25 - 13)^2)} = 101,838$$

$$d_4 = \sqrt{((0 - 0)^2 + (0 - 21)^2 + \dots + (5 - 13)^2)} = 31,702$$

$$d_5 = \sqrt{((0 - 0)^2 + (4 - 21)^2 + \dots + (0 - 13)^2)} = 34,395$$

$$d_{172} = \sqrt{((20 - 0)^2 + (9 - 21)^2 + \dots + (0 - 13)^2)} = 28,862$$

Tabel 4 menunjukkan hasil perhitungan jarak dari Rumus 1 terhadap data training dan data testing.

**Tabel 4.** Hasil Perhitungan Jarak

No.	Data uji 1	Data uji 2	Data uji 3	::	Data uji 43
1.	408,137	428,943	294,810	::	430,545
2.	33,332	32,016	192,031	::	9,055
3.	101,838	121,445	88,544	::	115,827
4.	31,702	27,000	191,541	::	9,165
5.	34,395	35,567	187,499	::	14,491
::	::	::	::	::	::
172.	28,862	51,039	158,606	::	34,264

2. Urutkan Jarak

Setelah perhitungan jarak selesai, maka selanjutnya mencari jarak  $k = 3$  dari hasil perhitungan euclidean distance untuk menemukan jarak terkecil. Berikut hasil pengurutan pada Tabel 5.

**Tabel 5.** Hasil Urutan K=3

Data uji 1	Data uji 2	Data uji 3	::	Data uji 43
3,5	3,6666667	49,833333	::	1
6	4,8333333	71,666667	::	1,6666667
12	6,8333333	74,333333	::	2,8333333

3. Rata-Rata Hasil Jarak Terkecil

Setelah diurutkan dengan mengambil 3 jarak terkecil dari masing-masing. Langkah berikutnya adalah menghitung nilai rata-rata dari hasil nilai dan kemudian membaginya dengan nilai  $k = 3$ , sehingga data uji 1 =  $(3,5 + 6 + 12) / 43 = 7,17$ . Berikut nilai aktual dan nilai prediksi penjualan parfum dari 43 data uji.

**Tabel 6.** Hasil Prediksi Penjualan Parfum

Data Uji	Nilai Aktual	Nilai Prediksi
1	10,67	7,17
2	5,33	5,11
3	63,50	65,28
4	35,83	39,67
5	28,00	25,28
::	::	::
43	2,50	1,83

**3.2. Implementasi Python**

**3.2.1 Input Data**

Langkah pertama dalam analisis data menggunakan bahasa pemrograman Python adalah mengimpor data ke Jupyter Notebook. Library Pandas diperlukan untuk melakukan input data dan analisis data. Setelah selesai, maka dapat ditampilkan dataset penjualan parfum sebanyak 215 data. Gambar 3 menunjukkan tampilan dataset penjualan parfum di Python.

```
[50]: #Import library
import numpy as np
import pandas as pd

[504]: #Input dataset
dataset = pd.read_excel('Dataset Penjualan Parfum.xlsx') #('Dataset K-NN Regresi.xlsx')

[528]: #menampilkan dataset, 5 data dari atas
dataset.head()

[525]:
```

No	Nama Parfum	Januari	Februari	Maret	April	Mei	Juni	Label
0	1 AIGNER BLACK MAN	71	245	211	185	182	116	168.333333
1	2 VANILLA CHOCO	0	0	0	0	1	0	0.1666667
2	3 AIGNER DEBUT	41	37	25	84	63	25	45.833333
3	4 AIGNER IN LEATHER M	0	0	0	0	0	5	0.833333
4	5 AIGNER STARLIGHT	0	4	0	0	15	0	3.166667

**Gambar 3.** Tampilan Dataset Teratas

### 3.2.2 Data Splitting

Setelah memasukkan dataset, kemudian memecah data. Namun, pemrograman ini belum menentukan atribut dan labelnya, jadi terlebih dahulu dilakukan penentuan atribut, yaitu penjualan dari bulan 1 hingga 6, dan labelnya adalah rata-rata penjualan parfum bulan 1 hingga 6. Selanjutnya, data training dan data testing dipisahkan. Pada penelitian ini, 20% data testing dan 80% data training dibagi.

### 3.2.3 Training K-Nearest Neighbor

Training K-Nearest Neighbor dilakukan di Jupyter Notebook menggunakan bahasa pemrograman Python. Proses K-Nearest Neighbor dilakukan dengan menambahkan fungsi Scikit-learn neighbors, menambahkan nilai k = 3, dan juga perhitungan jarak euclidean distance. Dapat dilihat Gambar 4 proses training K-NN di pemrograman Python.

```
[530]: #Training K-NN Model

from sklearn.neighbors import KNeighborsRegressor

k = 3
model = KNeighborsRegressor(n_neighbors=k, metric='euclidean')
model.fit(x_train, y_train)

[530]: + KNeighborsRegressor
KNeighborsRegressor(metric='euclidean', n_neighbors=3)
```

**Gambar 4.** Training K-NN di Python

Setelah Training K-Nearest Neighbor selesai, selanjutnya dilakukan prediksi terhadap model K-NN yang sudah di training. Gambar 5 menunjukkan hasil prediksi penjualan parfum.

```
Nilai aktual :
[ 10.66666667  5.33333333  63.5          35.83333333  28.
 164.83333333  10.5         30.33333333  5.83333333  12.
  7.5          7.5          38.          2.5          21.
 70.16666667  20.         41.16666667  3.          37.33333333
  2.16666667  9.16666667  113.         1.66666667  1.66666667
 28.33333333  53.5        2.16666667  23.         16.83333333
 13.          14.66666667  8.66666667  8.5         28.33333333
 11.83333333  6.16666667  30.          88.83333333  47.83333333
  4.16666667  28.83333333  2.5          ]

Nilai prediksi :
[  7.16666667  5.11111111  65.27777778  39.66666667  25.27777778
 153.33333333  9.          28.27777778  5.5         10.77777778
  3.88888889  3.66666667  35.16666667  2.66666667  13.05555556
 63.61111111  10.83333333  47.61111111  2.94444444  29.5
  2.88888889  6.66666667  122.88888889  2.16666667  1.44444444
 18.33333333  48.44444444  1.44444444  22.5         18.5
 14.22222222  16.61111111  4.33333333  8.22222222  25.11111111
 11.5         7.44444444  22.94444444  78.72222222  43.5
  2.11111111  30.22222222  1.83333333]
```

**Gambar 5.** Hasil Prediksi

Aigner Blue Emotion adalah parfum terlaris dengan total 153 ml penjualan pada bulan berikutnya, seperti yang ditunjukkan oleh hasil prediksi penjualan parfum pada Gambar 5.

### 3.2.4 Evaluasi Model

Evaluasi model terhadap kinerja dari algoritma K-NN pada penelitian ini menggunakan Mean Square Error (MSE). MSE adalah perhitungan evaluasi model terhadap kesalahan peramalan. Melakukan evaluasi model dengan MSE dapat menunjukkan seberapa besar perbedaan antara hasil estimasi dan yang akan diestimasi [17]. Nilai MSE yang lebih rendah menunjukkan kinerja prediksi yang lebih baik [18]. Berdasarkan hasil prediksi yang telah ditampilkan pada Gambar 5 untuk mengevaluasi akurasi dari model K-NN dengan perhitungan nilai MSE dengan rumus 2.

$$MSE = \frac{\sum(y-y')^2}{n} \tag{2}$$

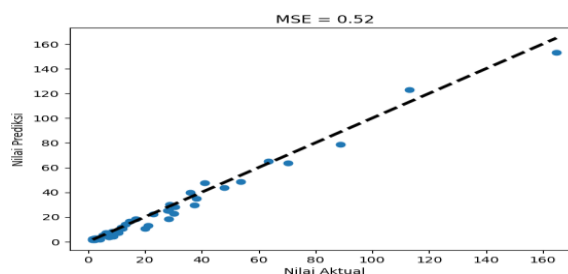
dimana y adalah label aktual, y' adalah label prediksi, dan n adalah jumlah data. Tabel 7 menunjukkan perhitungan evaluasi model dengan MSE..

**Tabel 7.** Hasil Evaluasi MSE

Data Uji	Label Asli	Label Prediksi	((Asli-Prediksi)^2)/n
1	10,67	7,17	0,2849
2	5,33	5,11	0,0011
3	63,50	65,28	0,0735

Data Uji	Label Asli	Label Prediksi	$((\text{Asli}-\text{Prediksi})^2)/n$
4	35,83	39,67	0,3417
5	28,00	25,28	0,1723
6	164,83	153,33	3,0756
7	10,50	9,00	0,0523
8	30,33	28,28	0,0983
9	5,83	5,50	0,0026
10	12,00	10,78	0,0347
11	7,50	3,89	0,3033
12	7,50	3,67	0,3417
13	38,00	35,17	0,1867
14	2,50	2,67	0,0006
15	21,00	13,06	1,4678
16	70,17	63,61	0,9994
17	20,00	10,83	1,9541
18	41,17	47,61	0,9658
19	3,00	2,94	0,0001
20	37,33	29,50	1,4270
21	2,17	2,89	0,0121
22	9,17	6,67	0,1453
23	113,00	122,89	2,2742
24	1,67	2,17	0,0058
25	1,67	1,44	0,0011
26	28,33	18,33	2,3256
27	53,50	48,44	0,5944
28	2,17	1,44	0,0121
29	23,00	22,50	0,0058
30	16,83	18,50	0,0646
31	13,00	14,22	0,0347
32	14,67	16,61	0,0879
33	8,67	4,33	0,4367
34	8,50	8,22	0,0018
35	28,33	25,11	0,2415
36	11,83	11,50	0,0026
37	6,17	7,44	0,0380
38	30,00	22,94	1,1577
39	88,83	78,72	2,3775
40	47,83	43,50	0,4367
41	4,17	2,11	0,0983
42	28,83	30,22	0,0449
43	2,50	1,83	0,0103
Rata-rata keseluruhan MSE			0,52

Prediksi penjualan parfum dengan metode K-NN, hasil rata-rata evaluasi menggunakan MSE adalah 0,52, yang menunjukkan bahwa hasilnya sangat baik. Ini disebabkan oleh fakta bahwa hasil perhitungan semakin dekat dengan nilai aktual semakin rendah nilai MSE. Selanjutnya menampilkan grafik di Python dari hasil evaluasi prediksi penjualan parfum menggunakan metode K-NN dengan MSE. Gambar 6 menggambarkan hasil MSE dari penelitian ini.



**Gambar 6.** Grafik Evaluasi MSE

Gambar 6 menunjukkan bahwa titik-titik yang ada, dekat dan tidak ada yang jauh dari garis diagonal yang mengindikasikan perhitungan model mendekati nilai sebenarnya. Ini menunjukkan bahwa hasil prediksi memiliki kualitas yang baik.

## 4. KESIMPULAN

Prediksi penjualan parfum terlaris menggunakan metode K-NN, menggunakan jarak euclidean distance dengan parameter  $k = 3$  dapat digunakan. Aigner Blue Emotion adalah penjualan parfum terlaris dengan total penjualan 153 ml. Dari pengujian rata-rata evaluasi menggunakan MSE diperoleh hasil sebesar 0,52, yang menunjukkan bahwa hasilnya sangat baik. Ini disebabkan oleh fakta bahwa hasil perhitungan semakin dekat dengan nilai aktual semakin rendah nilai MSE.

## REFERENCES

- [1] Y. Handayani, T. Hidayat, and H. Arruhama, "Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Produk Terlaris Pada Toko Indah Jaya," *J. Tek. Inform. dan Desain Komun. Vis.*, vol. 2, no. 2, pp. 82–91, 2023.
- [2] D. Prasetyawan and R. Gatra, "Algoritma K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 56–67, 2022, doi: 10.14421/jiska.2022.7.1.56-67.
- [3] S. Abody, E. R. Br Gultom, S. Ramadhany, and A. Afifudin, "Prediksi Penjualan Sparepart Mobil Terlaris Menggunakan Metode K-Nearest Neighbor," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 6, p. 2003, 2022, doi: 10.30865/jurikom.v9i6.5189.
- [4] U. Hidayat, R. Helilintar, and ..., "Rekomendasi Restock Parfum Menggunakan Metode K-Means," *Pros. SEMNAS ...*, pp. 188–193, 2022, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/2582>
- [5] M. Fajri and D. Vionanda, "Prediksi Harga Minyak Mentah Dunia Menggunakan Metode Ensemble -Nearest Neighbor," vol. 7, pp. 17357–17368, 2023.
- [6] T. Nasution, "Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Kelulusan Mahasiswa Tepat Waktu," *J. Perangkat Lunak*, vol. 2, no. 1, pp. 1–14, 2020, doi: 10.32520/jupel.v2i1.944.
- [7] H. Said, N. Matondang, H. Nurramdhani Irmada, and S. Informasi, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi Application of K-Nearest Neighbor Algorithm to Predict Consumable Water Quality," vol. 21, no. 2, pp. 256–267, 2022, [Online]. Available: [www.kaggle.com](http://www.kaggle.com)
- [8] N. Nurdiana and A. Algifari, "Studi Komparasi Algoritma ID3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *INFOTECH J.*, vol. 6, no. 2, pp. 18–23, 2020.
- [9] Muhammad Romzi and B. Kurniawan, "Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma," *JTIM J. Tek. Inform. Mahakarya*, vol. 03, no. 2, pp. 37–44, 2020.
- [10] M. S. Hasibuan and Suhardi, "Analisis Sentimen Kebijakan Vaksin Covid-19 Menggunakan SVM dan C4.5," *J. Tek. Elektro Dan Komput. TRIAC*, pp. 19–21, 2022.
- [11] H. Said, N. H. Matondang, and H. N. Irmada, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi," *Techno.Com*, vol. 21, no. 2, pp. 256–267, 2022, doi: 10.33633/tc.v21i2.5901.
- [12] M. Rosadi, D. Aulia Nurhasanah, and M. Siddik Hasibuan, "Clustering Panjang Ruas Jalan di BBPJJN Sumut Menggunakan Algoritma K-Means," *J. Comput. Sci. Informatics Eng.*, vol. 02, no. 1, pp. 29–38, 2023, doi: 10.55537/cosie.v2i1.567.
- [13] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [14] Iswanto, Y. M. Pranoto, and R. A. Harianto, "Penerapan algoritma klasifikasi untuk sistem rekomendasi buy sell di forex trading," *Fasilkom*, vol. 10, no. 2, pp. 152–158, 2020.
- [15] S. Sukanto, Y. Adriyani, and R. Aulia, "Prediksi Kelompok UKT Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *JUITA J. Inform.*, vol. 8, no. 1, p. 121, 2020, doi: 10.30595/juita.v8i1.6267.
- [16] I. Yolanda and H. Fahmi, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT . Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor," vol. 3, no. 3, pp. 9–15, 2021.
- [17] H. Rofiq, K. C. Pelangi, and Y. Lasena, "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes," *J. Manaj. Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 8–15, 2020, [Online]. Available: <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19417.pdf>
- [18] N. Iksan, Y. P. Putra, and E. D. Udayanti, "Regresi Linier Untuk Prediksi Permintaan Sparepart Sepeda Motor," *ITEJ (Information Technol. Eng. Journals)*, vol. 03, no. 02, pp. 2548–2157, 2018.