

## Penerapan Data Mining untuk Prediksi Penjualan Produk Skincare Menggunakan Metode *KNN* (*K-Nearest Neighbors*): Studi Kasus Klinik Ayume Beauty Care

Putri Yulia Amanda<sup>1</sup>, Berlian Aisyah Damayanti<sup>2</sup>, Alisyah Akbar Choirun<sup>3</sup>, Selvi Novita Sari<sup>4</sup>, Siti Armiyanti<sup>5</sup>, M. Mahaputra Hidayat<sup>6</sup>

<sup>1,2,3,4,5,6</sup> Department of Informatics Engineering, Bhayangkara Universitas Surabaya Jl. Ahmad Yani No.114, Ketintang, Kec Gayungan, Kota SBY, Jawa Timur 60231

Email: <sup>1</sup>putriyulia207@gmail.com, <sup>2</sup>berlianaisyaa@gmail.com, <sup>3</sup>alsyibrchn@gmail.com, <sup>4</sup>selvinovsa19@gmail.com, <sup>5</sup>armiyanti3012@gmail.com, <sup>6</sup>mahaputra@ubhara.ac.id  
Email Penulis Korespondensi: <sup>1</sup>putriyulia207@gmail.com

**Abstrak**—Prediksi penjualan produk skincare menjadi tantangan penting dalam industri kecantikan, terutama untuk membantu klinik seperti Ayume Beauty Care dalam menyusun strategi penjualan yang efektif dan mengoptimalkan manajemen stok. Penelitian ini bertujuan menerapkan data mining dengan metode K-Nearest Neighbors (KNN) untuk memprediksi penjualan produk skincare, dengan mengandalkan data historis penjualan, demografi pelanggan, serta informasi promosi dan musiman. Langkah awal meliputi pengumpulan data, pembersihan, dan transformasi data agar sesuai dengan persyaratan algoritma KNN. Model KNN kemudian dilatih menggunakan data historis untuk mengidentifikasi pola dan tren penjualan. Evaluasi dilakukan menggunakan metrik akurasi, precision, dan recall untuk mengukur performa prediksi. Hasil penelitian menunjukkan bahwa metode KNN dapat memberikan prediksi yang cukup akurat, sehingga memungkinkan klinik untuk merencanakan strategi penjualan yang lebih tepat sasaran dan efisien. Penelitian ini juga memberikan rekomendasi praktis bagi klinik dalam menentukan waktu dan jenis promosi serta menjaga ketersediaan stok produk sesuai permintaan yang diprediksi. Penerapan prediksi berbasis data ini diharapkan dapat meningkatkan efektivitas bisnis Klinik Ayume Beauty Care dan memberikan kepuasan lebih kepada pelanggan.

**Kata Kunci:** Data Mining, KNN, Prediksi Penjualan, Klinik Ayume Beauty Care, Akurasi

**Abstract**—Skincare product sales prediction is an important challenge in the beauty industry, especially to help clinics such as Ayume Beauty Care in developing effective sales strategies and optimizing stock management. This study aims to apply data mining with the K-Nearest Neighbors (KNN) method to predict skincare product sales, relying on historical sales data, customer demographics, and promotional and seasonal information. The initial steps include data collection, cleaning, and data transformation to meet the requirements of the KNN algorithm. The KNN model is then trained using historical data to identify sales patterns and trends. Evaluation is carried out using accuracy, precision, and recall metrics to measure prediction performance. The results of the study show that the KNN method can provide fairly accurate predictions, allowing clinics to plan more targeted and efficient sales strategies. This study also provides practical recommendations for clinics in determining the time and type of promotion and maintaining product stock availability according to predicted demand. The application of this data-based prediction is expected to increase the effectiveness of the Ayume Beauty Care Clinic business and provide greater satisfaction to customers.

**Keywords:** Data Mining, KNN, Sales Prediction, Ayume Beauty Care Clinic, Accuracy

### 1. PENDAHULUAN

Industri kecantikan, khususnya produk skincare, mengalami pertumbuhan pesat seiring meningkatnya kesadaran masyarakat akan pentingnya perawatan kulit. Dengan meningkatnya minat konsumen terhadap produk skincare, bisnis-bisnis di sektor ini, seperti klinik kecantikan dan perawatan kulit, menghadapi tantangan dalam memprediksi tren penjualan untuk memastikan ketersediaan produk dan mengoptimalkan strategi pemasaran. Dalam konteks ini, Klinik Ayume Beauty Care berusaha untuk memahami pola dan tren penjualan produknya guna memenuhi permintaan konsumen serta meningkatkan efisiensi operasional[1].

Penerapan teknologi Data Mining menjadi solusi yang potensial dalam menghadapi tantangan tersebut. Data mining adalah proses analisis data untuk menemukan pola atau informasi yang bermanfaat dari data historis, sehingga dapat membantu dalam pengambilan keputusan. Salah satu metode dalam data mining yang populer untuk melakukan prediksi adalah *K-Nearest Neighbors* (KNN). Metode KNN bekerja dengan mengklasifikasi data baru berdasarkan kemiripannya dengan data-data yang ada, sehingga dapat digunakan untuk memprediksi hasil atau kategori dari suatu data baru. Dalam konteks prediksi penjualan, metode KNN dapat membantu klinik dalam memproyeksikan jumlah penjualan produk skincare berdasarkan pola yang terdapat pada data penjualan sebelumnya[2].

Penelitian ini bertujuan untuk menerapkan metode KNN dalam memprediksi penjualan produk skincare di Klinik Ayume Beauty Care. Dengan adanya prediksi penjualan ini, Klinik Ayume Beauty Care dapat lebih baik dalam mengelola persediaan produk, menyusun strategi pemasaran, serta meningkatkan kualitas layanan kepada pelanggan. Hasil dari penelitian ini diharapkan dapat menjadi referensi dalam penerapan metode data mining untuk memprediksi penjualan di sektor kecantikan, khususnya dalam lingkup klinik kecantikan dan perawatan kulit[3],[4].

## 2. METODOLOGI PENELITIAN

### 2.1 KNN

Metode *K-Nearest Neighbors* (KNN) digunakan untuk memprediksi penjualan produk skincare berdasarkan prinsip kedekatan antar data historis penjualan. Dalam penelitian ini, KNN diterapkan untuk mengidentifikasi data penjualan sebelumnya yang memiliki kemiripan dengan pola penjualan terkini. Dengan melihat pola data yang mirip ini, KNN dapat memprediksi jumlah produk yang akan terjual dalam waktu tertentu, sehingga memungkinkan klinik untuk lebih efisien dalam mengelola persediaan[5].

Dalam *K-Nearest Neighbors* (KNN), terdapat beberapa metode yang digunakan untuk mengukur jarak antara data baru (data uji) dengan data lama (data pelatihan). Dua metode yang umum digunakan adalah Manhattan Distance (City Block Distance) dan Euclidean Distance. Euclidean Distance merupakan metode yang paling sering digunakan dalam pengukuran jarak karena kesederhanaannya dan kemampuannya untuk menghitung jarak lurus antara dua titik dalam ruang multidimensi[6]. Euclidean Distance dihitung sebagai akar kuadrat dari jumlah kuadrat perbedaan antara koordinat yang sesuai dari dua titik. Formula ini memastikan bahwa setiap dimensi dalam data diperhitungkan dengan proporsional, memberikan hasil yang intuitif dan mudah diinterpretasikan. Dengan menggunakan Euclidean Distance, kita dapat menentukan data mana yang paling dekat dengan data baru dan menggunakannya untuk melakukan klasifikasi yang akurat. Langkah-langkah untuk menghitung metode *K-Nearest Neighbor* antara lain[7]:

1. Menentukan Parameter K : Pilih nilai K yang tepat untuk digunakan dalam prediksi, biasanya melalui eksperimen atau validasi silang.
2. Menghitung Jarak : Hitung jarak antara setiap data latih dan data uji dengan menggunakan metrik jarak Euclidean. Rumus jarak Euclidean adalah:

$$euc = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

..

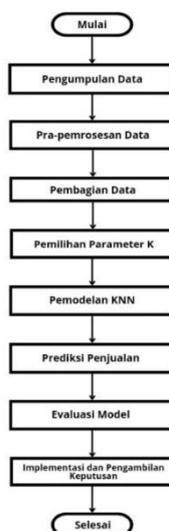
Keterangan :

- $p_i$  adalah sampel data dari data latih,
- $q_i$  adalah sampel data dari data uji,
- $n$  adalah dimensi data.

3. Mengurutkan Jarak : Urutkan jarak-jarak yang dihitung dari yang terkecil hingga yang terbesar.
4. Menentukan Jarak Terdekat Sampai Urutan K : Ambil K jarak terdekat dari data uji.
5. Memasangkan Kelas yang Bersesuaian : Identifikasi kelas dari K tetangga terdekat.
6. Mencari Jumlah Kelas : Hitung jumlah kelas dari tetangga terdekat dan tentukan kelas tersebut sebagai kelas prediksi untuk data yang dievaluasi

### Tahapan Penelitian

Penelitian Ini digambarkan menggunakan flowchart yang akan menggambarkan alur proses atau tahapan penelitian untuk memprediksi penjualan produk skincare di Klinik Ayume Beauty Care[8].



Gambar 1. Flowchart Tahapan Penelitian

Deskripsi :

1. Pengumpulan Data
  - a. Data Penjualan : Kumpulkan data historis penjualan produk skincare, termasuk tanggal, produk, jumlah terjual, dan informasi pelanggan (misalnya, usia, jenis kelamin).
  - b. Data Promosi: Catat informasi tentang promosi yang dilakukan (diskon, bundling produk, dll.).
  - c. Data Musiman: Pertimbangkan faktor musiman yang dapat mempengaruhi penjualan, seperti hari libur atau perayaan.
2. Pra-pemrosesan Data
  - a. Pembersihan Data: Hapus data yang tidak relevan atau duplikat.
  - b. Transformasi Data: Ubah data kategorikal menjadi numerik jika diperlukan (misalnya, menggunakan one-hot encoding untuk produk).
  - c. Normalisasi: Normalisasi fitur agar berada dalam rentang yang sama, karena KNN sensitif terhadap skala fitur.
3. Pembagian Data  
Training dan Testing: Bagi data menjadi set pelatihan (misalnya, 80%) dan set pengujian (20%). Ini penting untuk mengukur akurasi model.
4. Pemilihan Parameter K  
K-Nearest Neighbors: Tentukan nilai K (jumlah tetangga terdekat) yang optimal. Nilai K dapat diuji dengan menggunakan teknik validasi silang (cross-validation) untuk mendapatkan nilai K yang memberikan hasil terbaik.
5. Model KNN
  - a. Pelatihan Model: Gunakan data pelatihan untuk membangun model KNN. Model ini akan menghitung jarak antara titik data untuk menemukan tetangga terdekat berdasarkan fitur yang ada.
  - b. pelanggan, dan promosi yang akan datang.
6. Evaluasi Model
  - a. Akurasinya: Uji model dengan data pengujian untuk menilai akurasi prediksi. Gunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score untuk menilai kinerja model.
  - b. Analisis Hasil: Analisis hasil prediksi untuk memahami pola dan faktor yang mempengaruhi penjualan produk skincare.
7. Implementasi dan Pengambilan Keputusan
  - a. Rekomendasi Strategi: Berdasarkan hasil analisis, rekomendasikan strategi penjualan dan pemasaran yang lebih efektif, seperti promosi produk tertentu pada waktu-waktu tertentu berdasarkan prediksi penjualan.
  - b. Monitoring dan Pembaruan Model: Secara berkala perbarui model dengan data terbaru untuk meningkatkan akurasi dan relevansi prediksi.

## 2.1. Bahan Penelitian

Bahan Penelitian yang digunakan pada jurnal ini adalah Data pelanggan yang Diperoleh dari Data aktual penjualan bulan januari sampai dengan bulan desember[9].

**Tabel 1.** Data Stok Januari – April 2024 Klinik Ayume Beauty Care

No	Nama Barang	Januari	Februari	Maret	April
1	Acne Wash TTO	6	9	9	8
2	Facial Wash Bright	9	11	6	11
3	Toner Acne	3	1	0	1
4	Toner Lightening	3	10	2	4
5	Sunscreen Acne	4	4	2	2
6	Sunscreen Brightening	11	22	9	15
7	Skin Repair	18	19	19	32
8	Serum Brightening	0	1	0	1
9	Serum Acne	1	0	1	0
10	Serum Anti Aging	11	2	0	1
11	Premium Flek	0	3	4	2
12	Cream Malam	18	8	18	19
13	Acne Cream	6	4	3	3
14	Acne White	4	5	3	9
15	Whitening Cream	2	1	2	4
16	Gel Pagi C	1	2	0	3
17	Sunscreen Melasma	1	3	4	2
18	Hyalumoist	1	2	0	2
19	Anti Iritasi	1	13	4	8
20	Hair Growth Serum	1	1	2	1
21	CC Cream	0	0	1	1
22	Body White Pagi	1	0	1	0
23	Night Body Lotion	0	1	0	1
24	Bioxsine Shampoo	3	10	10	2

Tabel 2. Data Stok Mei– September 2024 Klinik Ayume Beauty Care

No	Nama Barang	Mei	Juni	Juli	Agustus	September
1	Acne Wash TTO	2	5	2	6	6
2	Facial Wash Bright	6	11	6	8	7
3	Toner Acne	0	2	2	1	0
4	Toner Lightening	3	4	4	4	2
5	Sunscren Acne	1	3	5	2	2
6	Sunscren Brightening	10	7	12	8	11
7	Skin Repair	17	28	24	25	28
8	Serum Brightening	1	0	1	1	1
9	Serum Acne	0	1	0	0	1
10	Serum Anti Aging	4	2	1	2	0
11	Premium Flek	5	4	2	4	0
12	Cream Malam	18	13	10	12	12
13	Acne Cream	7	5	4	3	2
14	Acne White	0	4	0	5	0
15	Whitening Cream	2	6	2	4	3
16	Gel Pagi C	1	1	3	0	0
17	Sunscren Melasma	2	4	3	3	4
18	Hyalumoist	1	1	1	1	1
19	Anti Iritasi	3	3	2	0	0
20	Hair Growth Serum	0	1	0	0	0
21	CC Cream	0	1	1	1	2
22	Body White Pagi	0	0	0	0	1
23	Night Body Lotion	0	0	1	0	0
24	Bioxsine Shampoo	14	5	4	2	2

## 3. HASIL DAN PEMBAHASAN

### 3.1 Membaca Data

Data yang digunakan dalam penelitian ini yaitu data penjualan produk skincare berdasarkan penjualan dari bulan Januari – September 2024 yang berjalan dari Ayume Beauty. Kemudian data tersebut akan di pre-processing untuk diolah dalam memprediksi penjualan produk pada bulan Oktober.

### 3.2 Preprocessing

Selanjutnya tahapan *pre-processing*. Pada tahap ini langkah-langkah yang dilakukan antara lain :

1. Menghapus kolom non-numerik

```
python Salin kode  
  
data_cleaned = df.drop(columns=["No", "Nama Barang"], errors="ignore")
```

Gambar 2. Menghapus kolom non-numerik

2. Membuat kolom target (bulan selanjutnya)

```
# 3. Membuat Kolom Target (Bulan Selanjutnya)  
# Target = Pergerakan penjualan dari bulan terakhir (September) ke bulan berikutnya (Oktober)  
data_cleaned["Target"] = data_cleaned["September"]  
  
# Menghapus baris terakhir karena Target-nya kosong  
data_cleaned = data_cleaned.dropna()
```

Gambar 3. Menghapus kolom target (bulan selanjutnya)

3. Mengatasi *missing values*

```
# Cek Missing Value  
print("\nCek Missing Value Sebelum Preprocessing:")  
print(data_cleaned.isnull().sum())  
  
# 2. Mengatasi Missing Value menggunakan SimpleImputer (mengganti dengan mean)  
imputer = SimpleImputer(strategy="mean")  
data_imputed = pd.DataFrame(imputer.fit_transform(data_cleaned), columns=data_cleaned.columns)  
  
# Cek Missing Value setelah imputasi  
print("\nCek Missing Value Setelah Preprocessing:")  
print(data_imputed.isnull().sum())
```

Gambar 4. Mengatasi missing values

Cek *missing values* setelah *pre-processing* :

```
Cek Missing Value Setelah Preprocessing:  
Januari      0  
Februari     0  
Maret        0  
April        0  
Mei          0  
Juni         0  
Juli         0  
Agustus     0  
September    0  
Target       0  
dtype: int64
```

Gambar 5. Hasil pre-processing

#### 4. Membagi Data dan Normalisasi Data

```
# 3. Membagi Data  
# Fitur: Data Januari hingga September  
X = data_imputed.iloc[:, :-1] # Semua kolom Januari - September  
y = data_imputed.iloc[:, -1] # Kolom terakhir (September) sebagai target awal  
  
# Normalisasi data menggunakan StandardScaler  
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(X)
```

Gambar 6. Membagi data dan normalisasi data

#### 5. Membuat Model KNN

```
# 4. Membuat Model KNN  
knn = KNeighborsRegressor(n_neighbors=3)  
knn.fit(X_scaled, y)
```

Gambar 7. Membuat Model KNN

#### 6. Membuat Data Prediksi

```
# 5. Membuat Data Prediksi untuk Oktober  
predictions = knn.predict(X_scaled)  
  
# Membulatkan hasil prediksi ke bilangan bulat  
predictions_rounded = predictions.round().astype(int)
```

Gambar 8. Membuat Data Prediksi

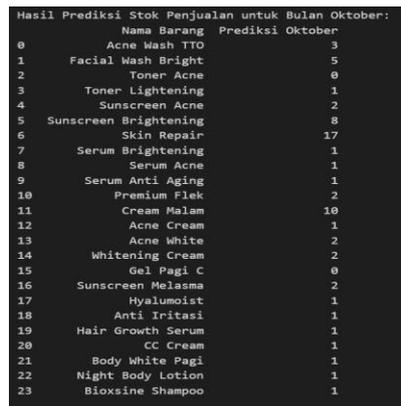
```
# Menyimpan index asli untuk referensi  
original_index = df.index  
print("Index Asli DataFrame:")  
print(original_index)
```

Gambar 9. Menyimpan index asli untuk referensi

#### 7. Menampilkan Hasil Prediksi

```
# Menambahkan hasil prediksi ke DataFrame asli  
df["Prediksi Oktober"] = pd.Series(predictions_rounded, index=original_index[:len(predictions_rounded)])  
  
# 7. Menampilkan Hasil  
print("Hasil Prediksi Stok Penjualan untuk Bulan Oktober:")  
print(df[["Nama Barang", "Prediksi Oktober"]])  
  
# Menyimpan hasil prediksi  
output_path = "prediksi_stok_oktober.xlsx"  
df.to_excel(output_path, index=False)  
print(f"\nHasil prediksi disimpan di: {output_path}")
```

Gambar 10. Menampilkan Hasil Prediksi



Hasil Prediksi Stok Penjualan untuk Bulan Oktober:		
	Nama Barang	Prediksi Oktober
0	Acne Wash TTO	3
1	Facial Wash Bright	5
2	Toner Acne	0
3	Toner Lightening	1
4	Sunscreen Acne	2
5	Sunscreen Brightening	8
6	Skin Repair	17
7	Serum Brightening	1
8	Serum Acne	1
9	Serum Anti Aging	1
10	Premium Flek	2
11	Cream Malam	10
12	Acne Cream	1
13	Acne White	2
14	Whitening Cream	2
15	Gel Pagi C	0
16	Sunscreen Melasma	2
17	Hyalumoist	1
18	Anti Iritasi	1
19	Hair Growth Serum	1
20	CC Cream	1
21	Body White Pagi	1
22	Night Body Lotion	1
23	Bioxsine Shampoo	1

**Gambar 10.** Hasil Prediksi Prediksi Stok Penjualan Bulan Oktober

Berdasarkan hasil dari Implementasi metode KNN diatas, dapat dilihat bahwa metode KNN ini cukup efektif dalam memprediksi produk skincare yang akan dibeli oleh konsumen dan juga mampu mengidentifikasi produk dengan potensi penjualan yang lebih tinggi sehingga dapat membantu dalam merencanakan strategi promosi yang lebih tepat pada sasaran.

## 8. KESIMPULAN

Penerapan Metode KNN (K-Nearest Neighbors) dalam prediksi penjualan produk Skincare di Klinik Ayume Beauty Care menunjukkan hasil yang menjanjikan. Dengan menggunakan model ini, Klinik Ayume Beauty Care dapat lebih efisien lagi dalam merencanakan produk serta menyusun strategi pemasaran yang tepat dan pengelolaan stok produk dengan lebih efisien. selain itu, penelitian ini bisa menjadi alat yang kuat untuk membantu bisnis pada bagian kecantikan untuk lebih terarah dan berbasis data.

Untuk mengoptimalkan penggunaan data mining dalam mendukung keputusan bisnis yang lebih baik di Klinik Ayume Beauty Care ini adalah yaitu dengan memastikan data yang digunakan lengkap dan berkualitas sehingga dapat membantu meningkatkan akurasi prediksi dan model pengembangan yang perlu diperbarui secara rutin dengan data terbaru untuk tetap relevan dengan perubahan tren pasar.

## REFERENCES

- [1] K. P. Sari and F. Firman, "Analisis Efektivitas Lembar Kerja dalam Meningkatkan Pemahaman Konsep Bangun Ruang Siswa SD," *Jurnal Pelita Ilmu Pendidikan*, vol. 1, no. 2, pp. 34–36, 2023.
- [2] H. Salma, "Hubungan antara Aktivitas Truk Batubara dan Konsentrasi Partikulat di Udara Provinsi Jambi," *Jurnal Kolaborasi Sains dan Ilmu Terapan*, vol. 2, 2023.
- [3] K. Tresia anatasya, Rikki Purba, "Kajian Jarak Aman Antara Arrival Traffic dan Departure Traffic Pada Unit Aerodrome Control Tower di Perum LPPNPI Kantor Cabang Medan," *Jurnal Kolaborasi Sains dan Ilmu Terapan*, vol. 1, no. 1, pp. 15–18.
- [4] I. M. Sianturi, "Perancangan Aplikasi Kompresi File Gambar Dengan Menggunakan Algoritma Stout Code," *Jurnal Pelita Ilmu Pendidikan*, vol. 2, no. 1, pp. 19–29, 2024.
- [5] W. Wati and F. A. Sianturi, "Implementasi Metode Topsis Dalam Merekomendasikan Pestisida Terbaik Pada Tanaman Padi Di Desa Rumbia," *Jurnal Sains dan Teknologi*, vol. 3, no. 2, pp. 31–35, 2022.
- [6] R. Purba and F. A. Sianturi, "Application of the C. 45 Algorithm in Measuring the Satisfaction Level of Hotel Visitors," *DISTANCE: Journal of Data Science, Technology, and Computer Science*, vol. 1, no. 1, pp. 18–26, 2021.
- [7] M. Fahmi and F. Sianturi, "Analisa algoritma Apriori pada pemesanan konsumen di café the l. Co coffe," *Jurnal sains dan teknologi*, vol. 1, no. 1, pp. 52–57, 2019.
- [8] F. A. Sianturi, "Analisa metode teorema bayes dalam mendiagnosa keguguran pada ibu hamil berdasarkan jenis makanan," *Jurnal Tekinkom (Teknik Informasi dan Komputer)*, vol. 2, no. 1, pp. 87–92, 2019.
- [9] F. A. Sianturi, P. M. Hasugian, A. Simangunsong, and B. Nadeak, "DATA MINING: Teori dan Aplikasi Weka." IOCS Publisher, 2019.