

# Implementasi Data Mining Untuk Prediksi Penjualan Produk Terlaris Di Alfamart Panglima Polim II Menggunakan Metode *K-Nearest Neighbor*

Dendi<sup>1</sup>, Budi Apriyanto<sup>2</sup>

<sup>1,2</sup>Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Indonesia

Email : [1dendidandi10@gmail.com](mailto:dendidandi10@gmail.com), [2dosen00628@unpam.ac.id](mailto:dosen00628@unpam.ac.id)

**Abstrak** - Penelitian ini bertujuan untuk mengimplementasikan data mining menggunakan metode K-Nearest Neighbor dalam memprediksi penjualan produk terlaris di Alfamart Panglima Polim II. Saat ini, kepala toko hanya mengandalkan analisis subjektif berdasarkan data penjualan bulan sebelumnya, yang tidak efektif dan dapat menyebabkan akumulasi stok barang yang berlebihan. Oleh karena itu, penelitian ini mengusulkan penggunaan data mining dan algoritma K-Nearest Neighbor sebagai pendekatan yang lebih efektif dalam menentukan produk yang banyak dibeli oleh konsumen. Data mining akan digunakan untuk menganalisis pola penjualan sebelumnya dan mengidentifikasi hubungan antara data penjualan yang ada. Metode K-Nearest Neighbor akan diterapkan untuk mengklasifikasikan produk berdasarkan kesamaan dengan data pembelajaran yang jaraknya terdekat. Diharapkan hasil penelitian ini dapat memberikan dukungan yang kuat kepada kepala toko dalam mengambil keputusan yang lebih akurat dalam menentukan produk terlaris. Selain itu, dengan mengimplementasikan data mining dan metode K-Nearest Neighbor, penelitian ini diharapkan dapat mengatasi masalah yang ada dalam analisis subjektif yang dilakukan saat ini. Dengan demikian, perusahaan dapat mengoptimalkan pengelolaan persediaan dan meningkatkan efisiensi operasional mereka.

**Kata kunci** : *K-Nearest Neighbor*, Produk Terlaris, Sistem Pendukung Keputusan.

**Abstract** - This research aims to implement data mining using the K-Nearest Neighbor method to predict the best-selling products at Alfamart Panglima Polim II. Currently, the store manager relies on subjective analysis based on previous month's sales data, which is ineffective and can lead to excessive stock accumulation. Therefore, this research proposes the use of data mining and the K-nearest neighbor algorithm as a more effective approach to determine products that are frequently purchased by consumers. Data mining will be used to analyze previous sales patterns and identify relationships among the existing sales data. The K-Nearest Neighbor method will be applied to classify products based on their similarity to the nearest neighbor in the training data. The expected outcome of this research is to provide strong support to the store manager in making more accurate decisions regarding the best-selling products. By implementing data mining and the K-Nearest Neighbor method, this research aims to address the limitations of the current subjective analysis approach. Consequently, the company can optimize inventory management and improve operational efficiency.

**Keywords** : *K-Nearest Neighbor*, Best Selling Products, Decision Support System

## 1. PENDAHULUAN

Dikutip dari (Pohan, Dar, & Irmayanti, 2022) prakiraan atau prediksi (forecasting) adalah perhitungan yang digunakan untuk meramalkan keadaan prospektif yang akan datang dengan menganalisis keadaan yang telah terjadi untuk dijadikan sebagai bahan pertimbangan dan sebagai bahan pengujian. Dalam memprediksi penjualan prospektif dapat menentukan perkiraan besarnya jumlah volume penjualan yang diharapkan, dan dapat memprediksi kemungkinan penjualan dimasa yang akan datang. Pemanfaatan prediksi diharapkan dapat mendukung seorang supervisor toko dalam mengambil sebuah keputusan untuk menentukan jumlah barang yang akan dipesan kepada supplier. Prakiraan atau prediksi berperan agar perusahaan dapat merencanakan persediaan barang yang akan dijual, oleh karena itu prakiraan dalam penjualan bertujuan agar dapat meminimalisir kemungkinan terjadinya kesalahan dalam penentuan jumlah barang yang akan distok.

Pada umumnya alfamart menjual berbagai produk mulai dari makanan ringan, minuman dan barang kebutuhan sehari-hari perusahaan ini bergerak pada bidang distribusi berbagai produk eceran yang pengoperasiannya telah memanfaatkan teknologi akan tetapi dalam proses prediksi penjualan produk yang banyak dibeli oleh konsumen kepala toko hanya menganalisa data yang bersifat subjektif berdasarkan pencapaian penjualan pada bulan sebelumnya yang mana hal ini sangat tidak efektif dan dapat berdampak negatif terhadap perusahaan seperti menumpuknya stok barang

digudang untuk itu diperlukan data mining yang kemudian di analisa menggunakan machine learning. Selain itu kepala toko tidak mempunyai pendukung keputusan yang kuat dalam menentukan produk terlaris sehingga penentuan produk terlaris masih sering terjadi error untuk itu algoritma metode k-nearest neighbor dapat dimanfaatkan dalam menentukan produk yang banyak di beli oleh konsumen. Berdasarkan permasalahan yang ada penulis mengusulkan penelitian dengan judul “ Implementasi Data Mining Untuk Prediksi Penjualan Produk Terlaris Di Alfamart Panglima Polim II Menggunakan Metode K-Nearest Neighbor ” diharapkan setelah penelitian ini dilakukan dapat menyelesaikan permasalahan tersebut.

## 2. METODE PENELITIAN

### 2.1 Metode Perancangan Bangun

Perancangan merupakan salah satu hal yang penting dalam membuat program. Adapun tujuan dari perancangan adalah untuk memberi gambaran yang jelas lengkap kepada pemrogram dan ahli teknik yang terlibat. Perancangan harus berguna dan mudah dipahami sehingga mudah digunakan. Perancangan adalah Sebuah Proses untuk mendefinisikan sesuatu yang akan dikerjakan dengan menggunakan teknik yang bervariasi serta didalamnya melibatkan deskripsi mengenai arsitektur serta detail komponen dan juga keterbatasan yang akan dialami dalam proses pengerjaannya (Ilham et al., 2018).

### 2.2 Metode K-Nearest Neighbor

Dikutip dari (Kafil, 2019) Algoritma k-nearest neighbor adalah sebuah metode algoritma untuk membuat klasifikasi terhadap suatu objek berdasarkan dengan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Algoritma k-nearest neighbor (K-NN) mempunyai konsep ”learning by analogy” yang mana data learning di deskripsikan dengan atribut numerik n-dimensi. Setiap data learning merepresentasikan sebuah titik yang di tandai dengan “c” dalam ruang n-dimensi.

Algoritma KNN memiliki lima cara dalam mencari tetangga terdekat yaitu : jarak euclidean, jarak manhattan, jarak cosine, jarak correlation dan jarak hamming akan tetapi dalam penelitian ini penulis menggunakan jarak euclidean dengan rumus sebagai berikut :

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Ket :

$D(i, j)$  = Jarak data ke i ke pusat cluster j

$X_{ki}$  = Data ke i pada atribut data ke k

$X_{kj}$  = Titik pusat ke j pada atribut ke k

Langkah-langkah untuk menghitung algoritma k-nearest neighbor :

- Menentukan nilai k.
- Menghitung nilai normalisasi data dengan rumus sebagai berikut :

$$\frac{X - \min(x)}{\max(x) - \min(x)}$$

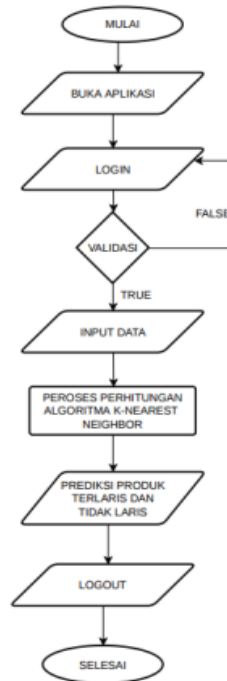
- Menghitung kuadrat jarak euclid (query instance) masing-masing objek terhadap training data yang diberikan.
- Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil.
- Mengumpulkan label class Y (klasifikasi k-nearest neighbor)

- f. Menggunakan kategori k-nearest neighbor yang paling mayoritas maka dapat diprediksikan nilai query instance yang telah dihitung.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Sistem Usulan

Sistem yang diusulkan pada penelitian ini dapat dilihat pada gambar dibawah ini:



**Gambar 1.** Sistem Usulan

Berdasarkan dari gambar sistem yang diusulkan pada penelitian ini, dapat dijelaskan Penentuan produk terlaris dengan sebuah sistem interface yang telah disediakan user dapat melakukan pengolahan data seperti input data, edit data dan menghapus data, setelah data lengkap maka sistem dapat menghitung algoritma *knearest neighbordan* menghasilkan output prediksi produk terlaris dan tidak laris

#### 3.2 Analisa Metode

Sebelum masuk ke tahapan perancangan sistem, terlebih dahulu dilakukan analisa pada data-data yang dijadikan sebagai sumber dari basis data pada perancangan sistem. Adapun sumber data yang dibutuhkan dalam perancangan sistem pada penelitian ini antara lain :

- a. Alternatif

Penelitian ini bertujuan untuk menentukan produk terlaris dan tidak terlaris di alfamart dalam beberapa tahun sebelumnya Adapun alternatif alternatif yang dimaksud adalah :

**Tabel 1.** Data Produk

No.	NAMA PRODUK	2017	2018	2019	2020	2021	LABEL
1	Fruit Tea 500ml	304	271	494	383	406	Terlaris
2	Aqua 600ml	311	442	394	302	388	Terlaris
3	Roma Malkist	330	426	354	290	414	Terlaris
4	Ff Skm Pch 545ml	366	350	329	410	357	Terlaris
5	Hot Wheels	352	416	358	309	358	Terlaris

**Tabel 2.** Data Uji

<b>Nama Produk</b>	<b>Th. 2017</b>	<b>Th. 2018</b>	<b>Th. 2019</b>	<b>Th. 2020</b>	<b>Th. 2021</b>	<b>Label</b>
JF Sulfur Anti Acne 90Gr	124	345	90	176	86	-

Untuk dapat melakukan perhitungan data mining dengan algoritma *knearest* *neighbor* dibutuhkan beberapa langkah, adapun langkah-langkah tersebut adalah sebagai berikut :

- a. Menentukan nilai k. dalam penentuan nilai k tidak ada aturan baku akan tetapi penentuan nilai k secara umum ditentukan dengan jumlah ganjil, karna untuk menghindari munculnya jumlah jarak yang sama dalam proses pengklasifikasian, untuk itu dalam perhitungan yang dilakukan penulis menggunakan nilai k nya adalah 3 (tiga).
- b. Menghitung nilai normalisasi data testing dan data uji dengan rumus :

$$\text{nilai (x) – nilai MIN(x) / nilai MAX(x) – nilai MIN(x)}$$

1) Normalisasi data testing

N1 : Fruit Tea 500ml

$$\text{Th. 2017} = ( 304 - 222 ) / ( 391 - 222 ) = 0.4852$$

$$\text{Th. 2018} = ( 271 - 246 ) / ( 445 - 246 ) = 0.1256$$

$$\text{Th. 2019} = ( 494 - 215 ) / ( 494 - 215 ) = 1$$

$$\text{Th. 2020} = ( 383 - 224 ) / ( 410 - 224 ) = 0.8548$$

$$\text{Th. 2021} = ( 406 - 245 ) / ( 415 - 245 ) = 0.9471$$

N2 : Aqua 600ml

$$\text{Th. 2017} = ( 311 - 222 ) / ( 391 - 222 ) = 0.5266$$

$$\text{Th. 2018} = ( 442 - 246 ) / ( 445 - 246 ) = 0.9849$$

$$\text{Th. 2019} = ( 394 - 215 ) / ( 494 - 215 ) = 0.6416$$

$$\text{Th. 2020} = ( 302 - 224 ) / ( 410 - 224 ) = 0.4194$$

$$\text{Th. 2021} = ( 388 - 245 ) / ( 415 - 245 ) = 0.8412$$

N3 : Roma Malkist 115

$$\text{Th. 2017} = ( 330 - 222 ) / ( 391 - 222 ) = 0.6391$$

$$\text{Th. 2018} = ( 426 - 246 ) / ( 445 - 246 ) = 0.9045$$

$$\text{Th. 2019} = ( 354 - 215 ) / ( 494 - 215 ) = 0.4982$$

$$\text{Th. 2020} = ( 290 - 224 ) / ( 410 - 224 ) = 0.3548$$

$$\text{Th. 2021} = ( 414 - 245 ) / ( 415 - 245 ) = 0.9941$$

N4 : FF SKM PCH 545ML

$$\text{Th. 2017} = ( 366 - 222 ) / ( 391 - 222 ) = 0.8521$$

$$\text{Th. 2018} = ( 350 - 246 ) / ( 445 - 246 ) = 0.5226$$

$$\text{Th. 2019} = ( 329 - 215 ) / ( 494 - 215 ) = 0.4086$$

$$\text{Th. 2020} = ( 410 - 224 ) / ( 410 - 224 ) = 1$$

$$\text{Th. 2021} = (357 - 245) / (415 - 245) = 0.6588$$

N5 : Hot Wheels

$$\text{Th. 2017} = (352 - 222) / (391 - 222) = 0.7692$$

$$\text{Th. 2018} = (416 - 246) / (445 - 246) = 0.8543$$

$$\text{Th. 2019} = (358 - 215) / (494 - 215) = 0.5125$$

$$\text{Th. 2020} = (309 - 224) / (410 - 224) = 0.4570$$

$$\text{Th. 2021} = (358 - 245) / (415 - 245) = 0.6647$$

**Tabel 3.** Data Testing Normalisasi

N...	DATA TESTING NORMALISASI				
	Th. 2017	Th. 2018	Th. 2019	Th. 2020	Th. 2021
1	0.4852	0.1256	1.0000	0.8548	0.9471
2	0.5266	0.9849	0.6416	0.4194	0.8412
3	0.6391	0.9045	0.4982	0.3548	0.9941
4	0.8521	0.5226	0.4086	1.0000	0.6588
5	0.7692	0.8543	0.5125	0.4570	0.6647

Berikut hasil dari sebuah normalisasi dari permasing-masing item tersebut

2) Normalisasi data uji

N1 : JF Sulfur Anti Acne 90Gr

$$\text{Th. 2017} = (124 - 222) / (391 - 222) = -0.5799$$

$$\text{Th. 2018} = (345 - 246) / (445 - 246) = 0.4975$$

$$\text{Th. 2019} = (90 - 215) / (494 - 215) = -0.4480$$

$$\text{Th. 2020} = (176 - 224) / (410 - 224) = -0.2581$$

$$\text{Th. 2021} = (86 - 245) / (415 - 245) = -0.9353$$

**Tabel 4.** Data Uji Normalisasi

N...	DATA UJI NORMALISASI				
	Th. 2017	Th. 2018	Th. 2019	Th. 2020	Th. 2021
1	-0.5799	0.4975	-0.4480	-0.2581	-0.9353

c. Menghitung jarak antara data uji dan data testing dengan menggunakan perhitungan euclidean distance sebagai berikut.

d.....	NILAI EUCLIDEAN DISTANCE
D1	2,855
D2	2,503
D3	2,578
D4	2,628
D5	2,438

$$d_1 = \sqrt{(0,485 - (-0,579))^2 + (0,125 - 0,497)^2 + (1,000 - (-0,448))^2 + (0,854 - (-0,258))^2 + (0,947 - (-0,935))^2} = 2,855$$

$$d_2 = \sqrt{(0,526 - (-0,579))^2 + (0,984 - 0,497)^2 + (0,641 - (-0,448))^2 + (0,419 - (-0,258))^2 + (0,841 - (-0,935))^2} = 2,503$$

$$d_3 = \sqrt{(0,639 - (-0,579))^2 + (0,904 - 0,497)^2 + (0,498 - (-0,448))^2 + (0,354 - (-0,258))^2 + (0,994 - (-0,935))^2} = 2,578$$

$$d_4 = \sqrt{(0,852 - (-0,579))^2 + (0,522 - 0,497)^2 + (0,408 - (-0,448))^2 + (1,000 - (-0,258))^2 + (0,658 - (-0,935))^2} = 2,628$$

$$d_5 = \sqrt{(0,769 - (-0,579))^2 + (0,854 - 0,497)^2 + (0,512 - (-0,448))^2 + (0,457 - (-0,258))^2 + (0,664 - (-0,935))^2} = 2,438$$

- d. Menentukan kelompok hasil data uji berdasarkan jarak antar data terdekat dengan nilai k yang telah ditentukan sebelumnya adalah 3 maka diambil nilai terkecil dari euclidean distance yaitu d31, d32 dan d40

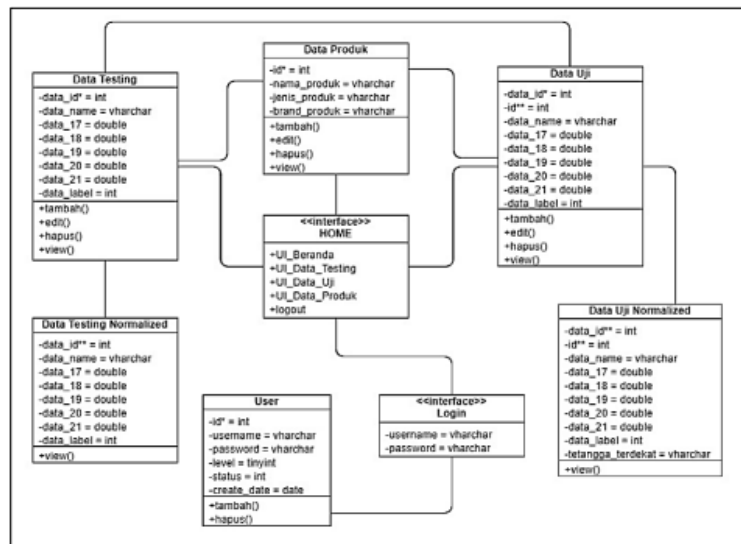
**Tabel 5.** Kelompok Data Hasil Uji

d.....	NAMA PRODUK	JARAK	LABEL
d31	Silverqueen Chunk	1,856	Tidak Terlaris
d32	Le Mineral 600ml	1,867	Tidak Terlaris
d40	Sweety Pampers	1,892	Tidak Terlaris

- e. Penggunaan metode k-nearest neighbor dapat disimpulkan bahwa untuk prediksi penjualan produk JF Sulfur Anti Acne 90Gr “Tidak Terlaris” bukan produk “terlaris”.

1. Perancangan Basis Data

Pada penelitian ini penulis akan menggunakan basis data atau database sebagai media penyimpanan hasil dari proses kompresi yang telah dilakukan. Basis data ini akan dibangun pada local host yang didukung oleh server apache. *Database* ini akan diberi nama knn, pada *database* ini akan dibuat delapan buah tabel dengan nama *user*, *home*, *log in*, *data testing*, *data testing normalized*, *data produk*, *data uji*, *data uji normalized*.



**Gambar 2.** Perancangan Class Diagram

2. Antarmuka Aplikasi

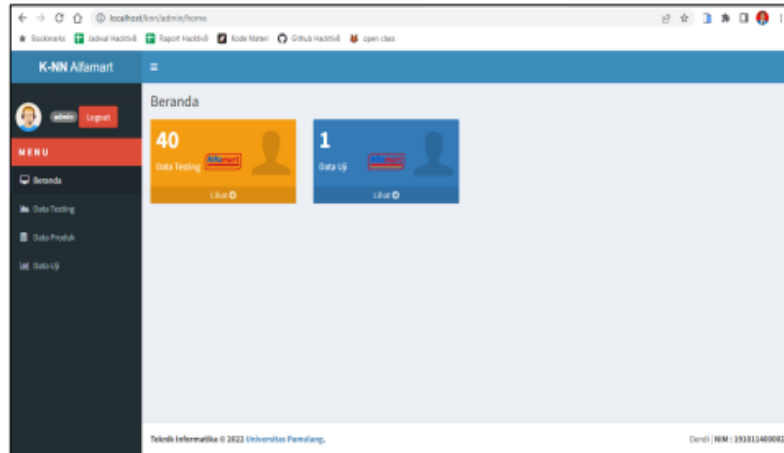
Berikut adalah implementasi dari perancangan antarmuka yang terdapat pada penelitian ini

a. Halaman Login



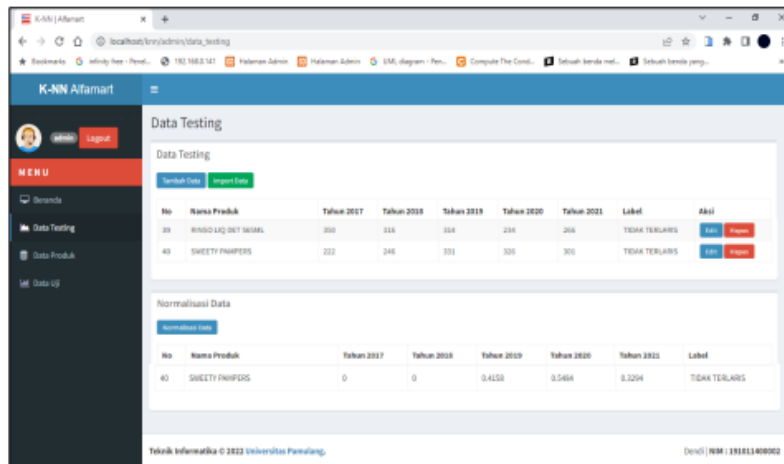
**Gambar 3.** Halaman Login

b. Halaman Menu Admin



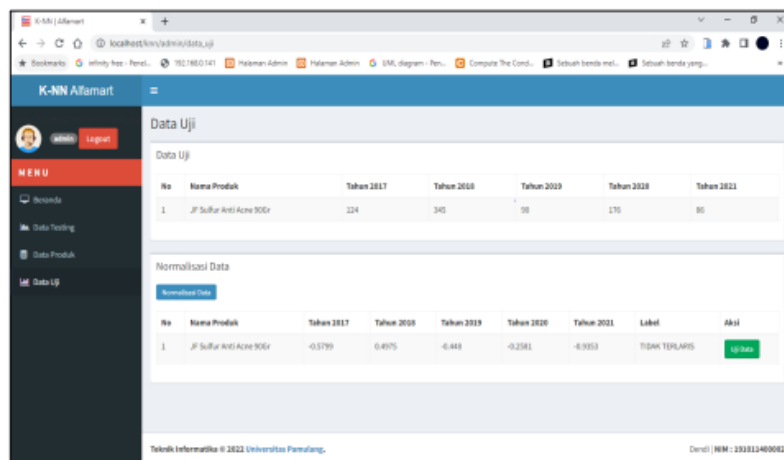
**Gambar 4.** Halaman Menu Admin

c. Halaman Utama Data Testing



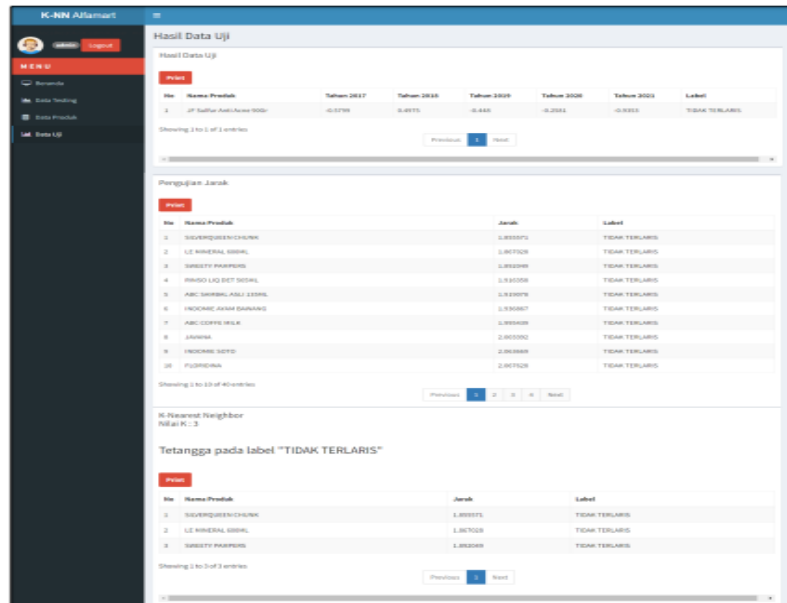
**Gambar 5.** Halaman Utama Data Testing

d. Halaman Menu Data Uji



**Gambar 6.** Halaman Menu Data Uji

e. Halaman Hasil Uji Produk



**Gambar 7.** Halaman Hasil Uji Produk

#### 4. KESIMPULAN

Berdasarkan uraian yang telah disampaikan penulis dalam bab pembahasan dan perancangan terhadap prediksi penjualan produk terlaris dan tidak terlaris di Alfa Mart Panglima Polim II dapat diambil kesimpulan sebagai berikut :

1. Sistem yang dibuat oleh penulis menerapkan machine learning dan dapat melakukan perhitungan secara cepat dengan memberi nilai inputan setiap variabel kemudian sistem akan melakukan perhitungan dan akan menghasilkan prediksi penjualan produk dengan tepat sama seperti perhitungan konvensional.
2. Algoritma metode k-nearest neighbor dapat diterapkan dalam aplikasi yang berbasis website dengan menggunakan bahasa pemrograman PHP dalam memprediksi penjualan produk sehingga diharapkan bisa membantu seorang kepala toko dalam menjawab sebuah permasalahan yang ada seperti menumpuknya stok barang digudang.
3. Penerapan machine learning dengan algoritma metode k-nearest neighbor dalam menentukan prediksi produk terlaris dan tidak terlaris menjadi lebih efektif dan akurat dengan melakukan perhitungan variabel inputan penjualan produk lima tahun terakhir (tahun : 2017, 2018, 2019, 2020, 2021) yaitu total penjualan setiap tahunnya.
4. Perhitungan yang dilakukan secara konvensional menunjukkan bahwa produk “JF Sulfur Anti Acne 90Gr” dengan nilai penjualan yang ada dan nilai k nya adalah tiga memiliki label “Tidak Terlaris” untuk mendukung keputusan yang adapenulis menguji data menggunakan aplikasi yang telah dibuat dan di terapkan dalam aplikasi berbasis website yang telah dibuat produk “JF Sulfur Anti Acne 90Gr” memiliki label yang sama seperti pada perhitungan konvensional.

#### 5. SARAN

Berdasarkan dari hasil analisa yang telah dilakukan dalam penelitian ini dengan berbagai keterbatasan yang dialami seperti masalah pemikiran dan waktu. Untuk itu saran dari penulis jika ingin mengembangkan sistem ini sebagai berikut :





- a. Penelitian selanjutnya penulis mengharapkan untuk dapat menggunakan metode yang lain dan metode yang terbaru agar dapat menjadi bahan perbandingan dalam penelitian selanjutnya.
- b. Pada objek penelitian diharapkan dapat menambahkan variabel yang akan diteliti sehingga keputusan dalam memprediksi jauh lebih akurat.
- c. Pembuatan sistem yang ada saat ini masih berbasis website disarankan untuk dapat membangun sistem yang berbasis mobile.

## **REFERENSI**

- Agusetiana, E., & Fitriani, A. S. (2022). Implementasi Data Mining Pada Pelanggan Telkom Menggunakan Metode K-Nearest Neighbor untuk Memprediksi Status Pelayanan. Seminar Nasional Inovasi Teknologi, 115-119.
- Agustin, A. P., Fauzan, A. C., & Harliana. (2022). Implementasi K-Nearest Neighbor Dengan Jarak Minkowski Untuk Deteksi Dini Covid-19 pada Citra Ct-Scan Paru-Paru. Jurnal Ilmiah Intech : Information Technology Journal of UMUS, 23-30.
- Aprilla C, D., Baskoro, D., Ambarwati, L., & Wicaksana, I. (2021). Belajar Data Mining dengan Rapid Miner. Semarang: Universitas Dian Nuswantoro.
- Dennis, A., Wixom, B. H., & Tegarden, D. (2015). Systems Analysis and Design : An Object-Oriented Approach with UML. New York: John Wiley & Sons Inc.