

## **PENERAPAN DATA MINING UNTUK PREDIKSI MEREK PAKAIAN YANG PALING DIMINATI DENGAN METODE K-NEAREST NEIGHBOR (STUDI KASUS : PT. MATAHARI DEPARTEMENT STORE BINJAI)**

**Andreas Pratama<sup>1</sup>, Budi Serasi Ginting<sup>2</sup>, Nurhayati<sup>3</sup>**  
STMIK Kaputama  
Email: [andreaspra14@gmail.com](mailto:andreaspra14@gmail.com)

### **ABSTRACT**

*One of the business activities that must be carried out to keep the company running and growing is sales. Decisions taken by corporate responsibility holders will affect the company in the future. One of the decisions that must be determined is the product to be sold for the next period. In determining the decision, a method is needed so that the decisions to be taken can be right on target. The technique used to predict the situation in the next period is called prediction. This study proposes the development of a clothing sales prediction application. The method used is the classification with the K-Nearest Neighbor algorithm. The results of data mining calculations using classification techniques with the K-Nearest Neighbor algorithm are the most predominant, it can be predicted that the number of clothing sales in the next period will increase with an average prediction of 14,900 per month and the most popular clothing brand is Cardinal.*

**Keywords: Prediction, Data Mining, K-Nearest Neighbor (KNN), Clothing**

### **PENDAHULUAN**

PT. Matahari Departement Store Tbk Binjai adalah industri ritel berskala besar, pengolahannya dibagi dan dipisah menjadi beberapa bagian departemen diantaranya pakaian dengan berbagai merek eksklusifnya. Perusahaan ini telah menggunakan sistem informasi penjualan untuk mendukung kinerja pihak perusahaan maupun untuk layanan informasi penjualan. Dilihat dari banyaknya permintaan konsumen akan produk pakaian, maka dibutuhkan prediksi untuk penjualan merek pakaian yang paling diminati oleh konsumen. Prediksi ini bertujuan untuk mempermudah manajemen perusahaan pada PT. Matahari Departement Store Tbk Binjai dalam melakukan perencanaan penyediaan stock pakaian serta dapat memberi informasi yang akurat pada perusahaan tentang merek pakaian yang paling diminati oleh konsumen.

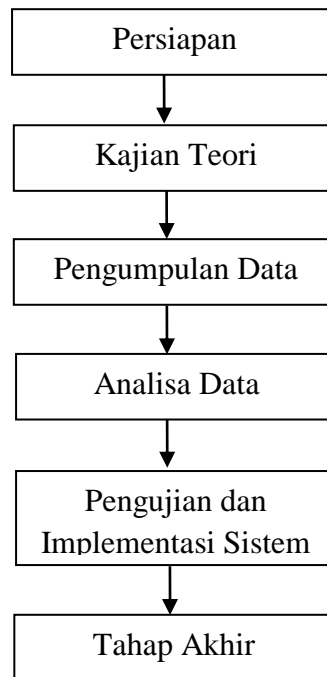
Prediksi atau *forecasting* penjualan merupakan seni dan ilmu yang dilakukan untuk mengetahui atau memperkirakan apa yang terjadi pada masa yang akan datang. Prediksi menjadi sangat penting karena penyusunan suatu rencana, diantaranya didasarkan pada suatu proyeksi atau ramalan. Oleh karena itu perusahaan perlu memprediksi apa yang akan terjadi pada masa yang akan datang yang digunakan sebagai landasan pembuat keputusan atau kebijakan untuk kelangsungan hidup perusahaan.

Namun kebutuhan informasi yang tinggi terkadang tidak diimbangi dengan penyajian informasi yang memadai, sehingga informasi tersebut masih harus digali ulang dari data yang jumlahnya sangat besar. Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan dalam memberikan informasi, tidak cukup hanya mengandalkan data operasional saja, sehingga diperlukan suatu analisis data agar dapat menggali potensi-potensi informasi yang tersedia.

## **METODE**

### **Metodelogi Penelitian**

Ada beberapa tahapan metode penelitian yang dilakukan dalam penyelesaian masalah. Tahapan tersebut sebagai berikut:



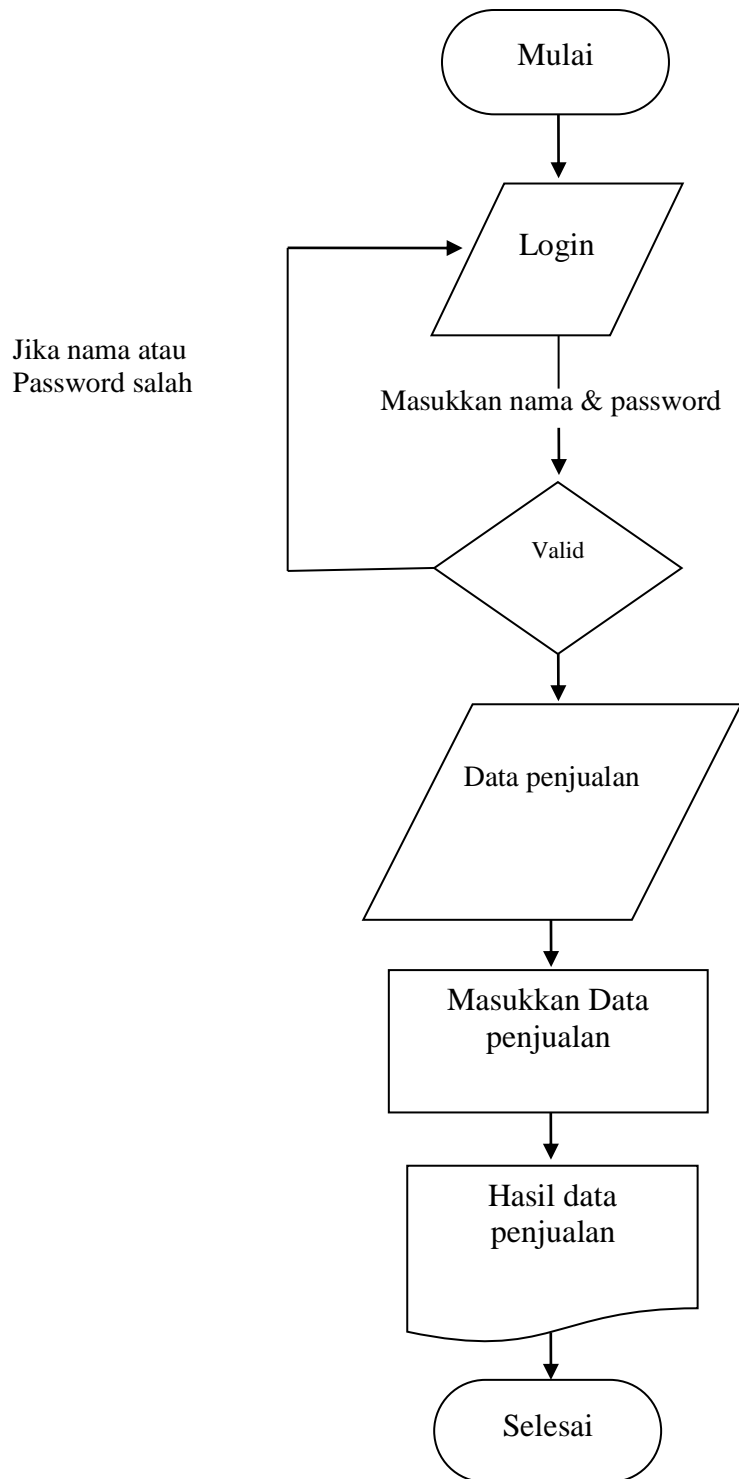
**Gambar 1. Metode Penelitian**

### **Analisis Program**

Berdasarkan pendahuluan dan mengacu pada landasan teori, dapat disusun suatu analisis sistem yang menjadi kerangka pemikiran penulis skripsi sebagai berikut:

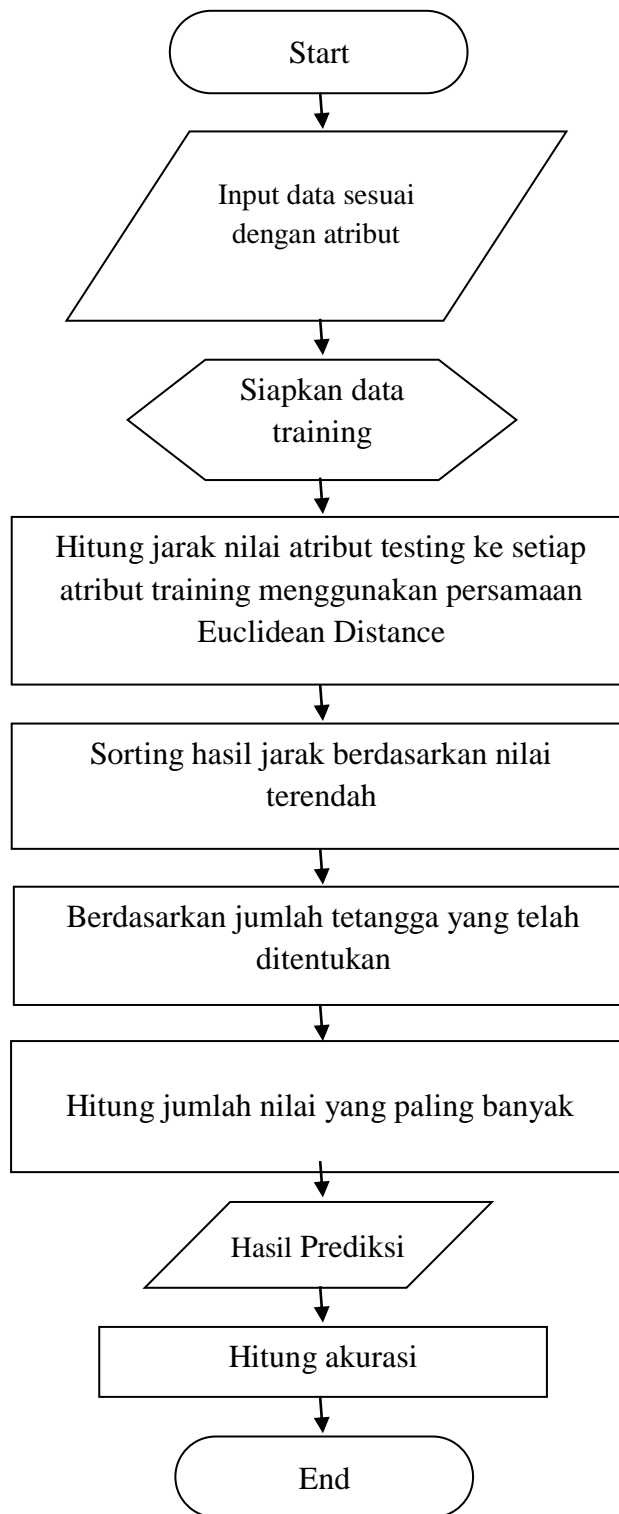
- 1) Dilakukan analisis sistem terhadap prediksi penjualan merek pakaian menggunakan metode klasifikasidengan algoritma *K-Nearest Neighbor* menggunakan *software visual basic*.
- 2) Menentukan kebutuhan sistem sehingga dapat dilakukan penganalisaan prediksi dari data penjualan pakaian menggunakan algoritma *K-Nearest Neighbor*.
- 3) Data penjualan pakaian merupakan data masukan, yang kemudian akan diproses dan dilatih dengan algoritma *K-Nearest Neighbor*, kemudian akan dilakukan pengujian terhadap data yang telah dilatih sebelumnya, setelah itu didapat hasil keluaran terhadap data yang telah diuji.

## Rancangan Flowchart Sistem



Gambar 2. Flowchart Sistem

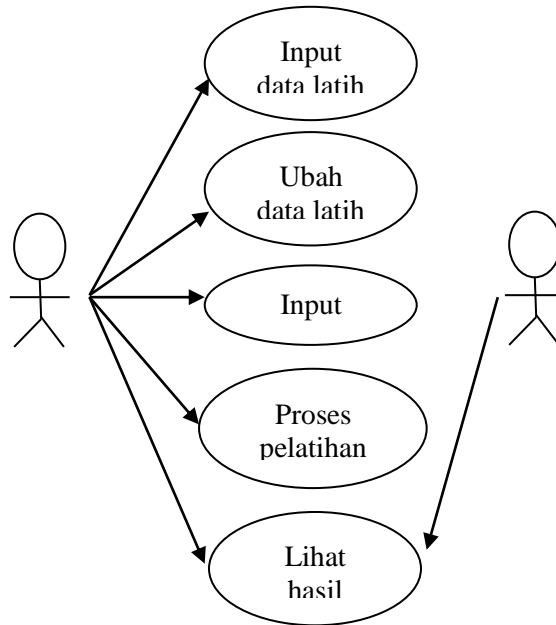
### Rancangan Flowchart K-Nearest Neighbor



Gambar 3. Flowchart Metode *K-Nearest Neighbor*

## Rancangan Flowchart Sistem

*Use case* diagram adalah suatu representasi atau model yang digunakan pada rekayasa perangkat lunak yang menunjukkan sekumpulan use case dan aktor serta hubungan diantara keduanya. Berikut gambar III.5 Rancangan *use case* sistem prediksi penjualan sepeda motor honda menggunakan metode *K-Nearest Neighbor*.



Gambar 4. Rancangan *Use Case*

## Rancangan Proses

### *Preprocessing*

*Preprocessing* merupakan tahap untuk mengelompokkan merek penjualan pakaian berdasarkan jumlah penjualan tiap bulan dan tahun untuk mempermudah dalam proses perhitungan prediksi. Setelah data terkelompokkan maka dapat dilihat jumlah penjualan untuk semua merek pakaian. Dapat dilihat pada beberapa tabel penjualan merek pakaian sebagai berikut :

Tabel 1. Sampel Data Cardinal dan Sampel Data Levis

Bln	2016	2017	2018	2019	Bln	2016	2017	2018	2019
Jan	4	7	4	2	Jan	0	0	0	2
Feb	6	6	7	4	Feb	0	1	5	0
Mar	4	4	2	1	Mar	2	1	1	1
Apr	2	3	4	2	Apr	1	1	1	1
Mei	7	11	2	5	Mei	3	0	1	0
Juni	10	2	4	3	Juni	3	2	0	1
Juli	7	8	8	8	Juli	5	0	2	0
Ag	4	6	7	2	Ag	2	2	0	1

Sep	12	4	2	4	Sep	2	0	1	2
Okt	8	8	4	6	Okt	1	1	1	3
Nov	9	5	3	6	Nov	1	3	2	1
Des	6	8	4	3	Des	0	2	1	0

### ***Transformation***

Pada tahap *transformation* ini hasil dari pengelompokkan data *preprocessing* kemudian digunakan untuk data *training*. Proses pembentukan data *training* berdasarkan data yang ada, data harus di seleksi terlebih dahulu untuk menentukan atribut mana yang dapat mempengaruhi penjualan sepeda motor yang paling diminati, data tersebut disebut data target. Dimana data target merupakan data yang berisikan atribut yang akan menjadi atribut yang relevan dan mendukung dalam prosese data mining. Adapun data *training* yang digunakan adalah sebagai berikut :

**Tabel 2. Data Training Penjualan Pakaian Cardinal**

No	Bln-1	Bln-2	Bln-3	Bln-4	Bln-5	Target
1	5	8	5	10	12	4
2	8	5	10	12	4	11
3	5	10	12	4	11	4
4	10	12	4	11	4	3
5	12	4	11	4	3	4
6	4	11	4	3	4	6
7	11	4	3	4	6	7
8	4	3	4	6	7	5
9	3	4	6	7	5	3
10	4	6	7	5	3	6
11	6	7	5	3	6	6
12	7	5	3	6	6	10
13	5	3	6	6	10	1
14	3	6	6	10	1	4
15	6	6	10	1	4	6
16	6	10	1	4	6	1
17	10	1	4	6	1	3
18	1	4	6	1	3	3
19	4	6	1	3	3	4
20	6	1	3	3	4	5
21	1	3	3	4	5	2
22	3	3	4	5	2	3
23	3	4	5	2	3	3
24	4	5	2	3	3	2
25	5	2	3	3	2	2
26	2	3	3	2	2	4
27	3	3	2	2	4	6
28	3	2	2	4	6	0
29	2	2	4	6	0	3
30	2	4	6	0	3	2
31	4	6	0	3	2	5

Berdasarkan dari tahapan data mining untuk algoritma *K-Nearest Neighbor*, adapun langkah-langkah dari *K-Nearest Neighbor*:

1. Penentuan nilai k. Penentuan nilai k yang digunakan tidak memiliki aturan yang baku, namun pada penelitian ini nilai k yang digunakan adalah 3.
2. Hitung jarak antar data *training* dan data uji (*test*) yang ada pada tahap *transformation* dengan menggunakan perhitungan *Euclidean Distance* sebagai berikut :

$$d_1 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 13,491$$

$$d_2 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 13,379$$

$$d_3 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 14,900$$

$$d_4 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 15,588$$

$$d_5 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 11,136$$

$$d_6 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 11,314$$

$$d_7 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 6,782$$

$$d_8 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 5,831$$

$$d_9 = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,681$$

$$d_{10} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 8,307$$

$$d_{11} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,416$$

$$d_{12} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 6,557$$

$$d_{13} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,746$$

$$d_{14} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 11,576$$

$$d_{15} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 9,327$$

$$d_{16} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 10,440$$

$$d_{17} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,071$$

$$d_{18} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,416$$

$$d_{19} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,000$$

$$d_{20} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 1,732$$

$$d_{21} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 6,164$$

$$d_{22} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 6,083$$

$$d_{23} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 5,745$$

$$d_{24} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 5,916$$

$$d_{25} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 3,873$$

$$d_{26} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 5,831$$

$$d_{27} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 4,472$$

$$d_{28} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 4,359$$

$$d_{29} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,874$$

$$d_{30} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,000$$

$$d_{31} = \sqrt{(5-6)^2+(8-0)^2+(5-3)^2+(10-2)^2+(12-5)^2} = 7,681$$

3. Pengurutan data hasil perhitungan. Jarak yang telah didapatkan kemudian diurutkan dari yang paling dekat jaraknya sampai yang paling jauh (*ascending*).

Setelah diurutkan diperoleh :

$d_{20} = 1,732$ ,  $d_{25} = 3,873$ ,  $d_{28} = 4,359$ ,  $d_{27} = 4,472$ ,  $d_{23} = 5,745$ ,  $d_{28} = 5,831$ ,  $d_{26} = 5,831$ ,  $d_{24} = 5,916$ ,  $d_{22} = 6,083$ ,  $d_{21} = 6,164$ ,  $d_{12} = 6,557$ ,  $d_7 = 6,782$ ,  $d_{19} = 7,000$ ,  $d_{30} = 7,000$ ,  $d_{17} = 7,071$ ,  $d_{11} = 7,416$ ,  $d_{18} = 7,416$ ,  $d_9 = 7,681$ ,  $d_{31} = 7,681$ ,  $d_{13} = 7,746$ ,  $d_{29} = 7,874$ ,  $d_{10} = 8,307$ ,  $d_{15} = 9,327$ ,  $d_{16} = 10,440$ ,  $d_5 = 11,136$ ,  $d_6 = 11,314$ ,  $d_{14} = 11,576$ ,  $d_2 = 13,379$ ,  $d_1 = 13,491$ ,  $d_3 = 14,900$ ,  $d_2 = 13,379$ .

4. Menentukan kelompok data hasil uji berdasarkan label mayoritas dari k tetangga terdekat. Karena nilai  $k = 3$  maka diambil jarak terkecil yaitu  $d_{20} = 1,732$ ,  $d_{25} = 3,873$ ,  $d_{28} = 4,359$ .
5. Dengan menggunakan kategori *K-Nearest Neighbor* yang paling mayoritas maka dapat diprediksikan jumlah penjualan pada periode berikutnya mengalami peningkatan dengan rata-rata prediksi setiap bulanya 14,900 dan di dapat merek pakaian yang paling diminati yaitu Cardinal .

## ANALISIS DAN PERANCANGAN

Dalam bab ini akan dijelaskan mengenai hasil uji coba perangkat lunak pembangunan algoritma *K-Nearest Neighbor* dengan menggunakan Visual Basic dan MySQL untuk mengetahui prediksi penjualan pada masa mendatang berdasarkan variabel tahun, bulan, produk dan jumlah penjualan yang paling dekat serta mengetahui merek pakaian mana saja yang paling banyak diminati.

Implementasi disini merupakan kegiatan dimana data-data yang sudah di transformasi diterapkan kedalam pemrograman yang digunakan dan diproses sesuai



metode klasifikasi dengan algoritma *K-Nearest Neighbor*. Sehingga dapat diketahui sejauh mana kinerja sistem dalam mengelola data dan dihasilkan informasi sesuai dengan kebutuhan pengguna. Selain itu dari hasil pengimplementasian yang dilakukan nantinya akan dilakukan analisa terhadap program atau *coding* sehingga dapat diketahui apakah sudah memenuhi tujuan yang ingin dicapai.

### **Pembahasan Antarmuka (*Interface*)**

Dalam pembahasan antarmuka ini akan dijelaskan mengenai hasil perancangan program menggunakan aplikasi Visual Basic dan MySQL yang dapat dilihat sebagai berikut :

#### **Menu Login**

Form Menu Login *Login* adalah tampilan *form* menu yang berisikan user *name* dan *password* yang harus diisi *user* ketika akan menggunakan sistem.



**Gambar 5. Menu Login**

#### **Form Menu Utama**

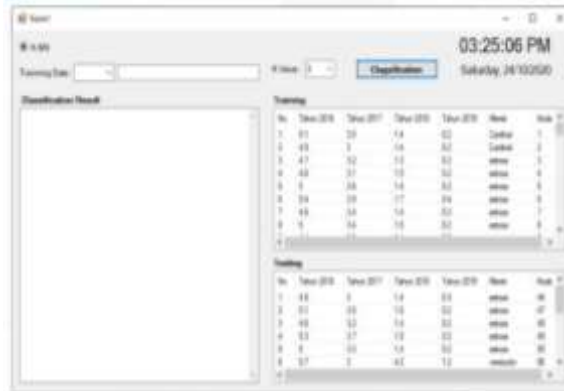
Saat pertama kali menjalankan sistem akan muncul tampilan menu utama. Pada tampilan menu utama terdapat menu open, proses dan logout.



**Gambar 6. FormMenu Utama**

## Form Menu Proses

Form Menu Proses adalah tampilan *form* menu yang berisikan tampilan proses *K-Nearest Neighbor*



**Gambar 7. Form Menu Proses**

## KESIMPULAN DAN SARAN

### Kesimpulan

Dalam penelitian laporan skripsi ini telah diuraikan bagaimana perancangan Penerapan Data Mining Untuk Prediksi Merek Pakaian Yang Paling Diminati Dengan Metode *K-Nearest Neighbor*” (Studi Kasus : PT. Matahari Departement Store Tbk Binjai). Adapun kesimpulan yang diperoleh adalah sebagai berikut :

1. Dengan adanya sistem prediksi yang diusulkan, maka PT. Matahari Departement Store Tbk Binjai memperoleh manfaat dalam perencanaan dan penjadwalan penyediaan stok pakaian untuk masa mendatang secara akurat.
2. *User* dapat melakukan proses penambahan, pengeditan, penghapusan data produk pakaian di dalam basis pengetahuan sehingga informasi dapat berkembang sesuai dengan kebutuhan dan dapat memberikan informasi yang tepat.
3. Dengan perhitungan data mining menggunakan teknik klasifikasi dengan algoritma *K-Nearest Neighbor* yang paling mayoritas maka dapat diprediksikan jumlah penjualan pada periode berikutnya mengalami peningkatan dengan rata-rata prediksi setiap bulanya 14,900 dan di dapat merek pakaian yang paling diminati yaitu Cardinal .

### Saran

Berdasarkan simpulan yang telah diuraikan, maka ada beberapa saran yang dapat disampaikan yaitu :

1. Diharapkan program sistem prediksi ini dapat diterapkan oleh PT. Matahari Departement Store Tbk Binjai agar mempermudah perusahaan dalam menentukan stok pakaian pada masa mendatang.
2. Pada penelitian yang akan datang dapat ditambahkan penerapan data mining dengan menggabungkan dua metode algoritma data mining sehingga akan menghasilkan hasil yang lebih variatif dan tentunya bernilai informasi yang lebih tinggi.

## DAFTAR PUSTAKA

- Amelia, M. winny, Lumenta, A. S. ., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika*. <https://doi.org/10.35793/jti.11.1.2017.17652>
- Amelia, Y. (2018). Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor. *Palembang: Universitas Islam Negeri Raden Fatah*.
- Fajar Astuti Hermawati, *Data Mining*, Andi, Yogyakarta, 2013
- Kamus Besar Bahasa Indonesia (KBBI), 2017.
- Hutami, R., & Astuti, E. Z. (2016). Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Furniture Pada CV. Octo Agung Jepara. *Universitas Dian Nuswantoro Semarang*.
- Kusrini. (2005). Proses Data Mining. *Mining of Massive Datasets*.
- Ladjamudin Al-Bahra. (2013). Analisis Dan Desain Sistem Informasi. In *Analisis Dan Desain Sistem Informasi*.
- Lastiansh, Sena. 2012. Pengertian User Interface. Jakarta: *PT. Elex Media Komputindo*.
- Mustakim, & Oktaviani, G. (2016). Algoritma K-Nearest Neighbor Classification. *Jurnal Sains, Teknologi Dan Industri*.
- Nugroho, B. (2005). Database Relasional dengan MySQL. In *Andi: Yogyakarta*.
- Nugroho, B. (2017). Dasar Pemrograman Web PHP-MySQL dengan Dreamweaver. In *Gava Media*.
- Prasetyo, E. (2014). DATA MINING Mengolah Data Menjadi Informasi Menggunakan Matlab. In *penerbit andi*.
- Sugiarti, Y. (2013). Pengantar UML (Unified Modeling Language). *Graha Ilmu*. <https://doi.org/http://dx.doi.org/10.1109/ICCSN.2010.19>
- Yatini B, I. (2014). Aplikasi pengolahan citra berbasis web menggunakan javascript dan jquery. *Jurnal Teknik*.
- Sutarman. 2007. *Membangun Aplikasi Web Dengan PHP dan MySQL*. Edisi Kedua, Cetakan Pertama, Graha Ilmu, Yogyakarta.
- Sutedjo Budi, *Perencanaan & Pembangunan Sistem Visual Basic*, Andi, Yogyakarta, 2009.
- Yatini, I. 2010. *Flowchat Algoritma dan Pemrograman Menggunakan Bahasa C++ Builder*. Edisi Pertama, Cetakan Pertama, Graha Ilmu, Yogyakarta.