

Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode *K-Nearest Neighbor*

Sri Puspita Dewi*, Nurwati, Elly Rahayu

Ilmu Komputer, Sistem Informasi, STMIK Royal Kisaran, Kisaran, Indonesia

Email: ^{1,*}sripuspitadewi074@gmail.com, ²nurwati763@gmail.com, ³ellyrahayu68@gmail.com

Email Penulis Korespondensi: sripuspitadewi074@gmail.com

Submitted: 16/03/2022; Accepted: 26/03/2022; Published: 31/03/2022

Abstrak—Penerapan Data Mining sangatlah dibutuhkan oleh UD Andar dikarenakan usaha dagang ini menjual berbagai macam jenis produk. Usaha Dagang ini menjual jamu serbuk, kantong plastik, bahan pembuat makanan & minuman, dan makanan frozen yang diminati oleh para konsumen. Dilihat dari banyaknya permintaan konsumen ternyata terdapat beberapa produk terlaris dan tidak terlaris, sehingga berdasarkan data 1 tahun terakhir, maka dibutuhkan sebuah prediksi penjualan produk terlaris, agar mempermudah pihak usaha dagang dalam perencanaan penyedia stok. Dikarenakan sistem yang sedang berjalan saat ini masih manual, untuk itu data yang didapat kurang akurat dan efisien. Maka untuk mengatasi hal ini, diperlukan sebuah sistem prediksi penjualan produk terlaris dengan teknik data mining yang menggunakan metode *k nearest neighbor*. Penelitian ini menghasilkan sebuah sistem algoritma *k nearest neighbor* pada teknik data mining yang membantu untuk memprediksi penjualan produk terlaris pada UD Andar.

Kata Kunci: Data Mining; K-Nearest Neighbor; Penjualan; Prediksi; Produk

Abstract—The implementation of Data Mining is very much needed by UD Andar because this trading business sells various types of products. This trading business sells powdered herbal medicine, plastic bags, food & beverage ingredients, and frozen foods that are in demand by consumers. Judging from the large number of consumer requests, it turns out that there are several best-selling and not-selling products, so based on the last 1 year data, a prediction of the best-selling product sales is needed, in order to make it easier for trading businesses in planning stock providers. Because the current system is still manual, the data obtained is less accurate and efficient. So to overcome this, we need a sales prediction system for best-selling products with data mining techniques using the *k nearest neighbor* method. This research produces a system of *k nearest neighbor* algorithms in data mining techniques that help to predict the sales of the best-selling products at UD Andar.

Keywords: Data Mining; K-Nearest Neighbor; Sales; Predictions; Products

1. PENDAHULUAN

Prediksi/peramalan (forecasting) merupakan suatu kegiatan meramalkan penjualan dimasa mendatang berarti menentukan perkiraan besarnya volume penjualan, bahkan menentukan potensi penjualan dan luas pasar yang dikuasai dimasa mendatang[1]. Selain itu membantu perusahaan dalam melakukan perencanaan penyediaan stok, karena prediksi ini memberikan *output* terbaik bagi perusahaan sehingga dapat meminimalisir kesalahan perencanaan dapat ditekankan seminimal mungkin.

Penjualan merupakan suatu kegiatan usaha yang dapat memberikan keuntungan lebih bagi sebuah usaha dagang retailer. UD Andar ini merupakan usaha dagang yang bergerak dibidang retailer. Produk yang dijual oleh UD Andar yaitu : Jamu serbuk, kantong plastik, bahan pembuat makanan & minuman, dan makanan frozen. Sistem yang sedang berjalan pada UD Andar masih manual yaitu belum adanya sistem yang dapat mengetahui produk mana saja yang paling diminati oleh konsumen serta belum tersedianya metode yang baik dalam penentuan produk terlaris pada UD Andar. Hal ini dikarenakan terjadinya peningkatan pada penjualan produk, yang disebabkan banyaknya permintaan para konsumen yang sangat beragam. Dengan ratusan produk yang terjual UD Andar mengalami kesulitan karena harus melakukan pengecekan struk belanja untuk menentukan produk mana saja yang terlaris berdasarkan data 1 (satu) tahun terakhir. Sehingga didapatkan permasalahan yang ada bahwa UD mengalami kesulitan dalam menentukan produk yang paling diminati oleh para konsumen. Maka dari itu, diperlukan sebuah sistem prediksi penjualan produk terlaris menggunakan metode *k nearest neighbor* untuk mengetahui produk apa saja yang paling diminati para konsumen.

Dalam hal ini Data Mining sangat berpengaruh dalam proses menggali sebuah data dari suatu kumpulan data yang berupa pengetahuan yang selama ini tidak diketahui secara manual [2]. Pada umumnya *Data mining* muncul dari banyaknya jumlah data yang tersimpan dalam suatu *database*. Data mining adalah pembelajaran mesin, pengenalan pola, database, statistik, dan teknik visualisasi yang digunakan untuk memecahkan masalah penggalian informasi dari repositori database besar[1]. *Data mining* sangat penting dalam proses penggalian data secara manual dari kumpulan data berupa pengetahuan yang tidak diketahui[2]. Berdasarkan pengertian *Data Mining* tersebut dapat disimpulkan bahwa *Data Mining* adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari *database* yang besar. Dalam melakukan peramalan begitu banyak metode yang bisa digunakan, salah satunya adalah metode *K Nearest Neighbor*. Menurut [2] metode *K-Nearest Neighbor* (K-NN) merupakan metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain, metode K-NN (*k nearest neighbor*) merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi. K-NN (*k nearest neighbor*) termasuk algoritma *supervised learning*, yang mana hasil dari *query instance* baru, diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN (*k nearest neighbor*). Kelas yang paling banyak muncul, yang akan menjadi kelas

hasil klasifikasi[3]. Tujuan dari algoritma K-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data[4]. Metode *K Nearest Neighbor* ini merupakan metode algoritma *matching learning* yang sangat sederhana dalam implementasinya. Dengan diterapkannya Algoritma *K Nearest Neighbor* dapat mempermudah UD Andar pada penjualan produk dengan mengambil data objek baru berdasarkan data yang letaknya terdekat dari data baru tersebut.

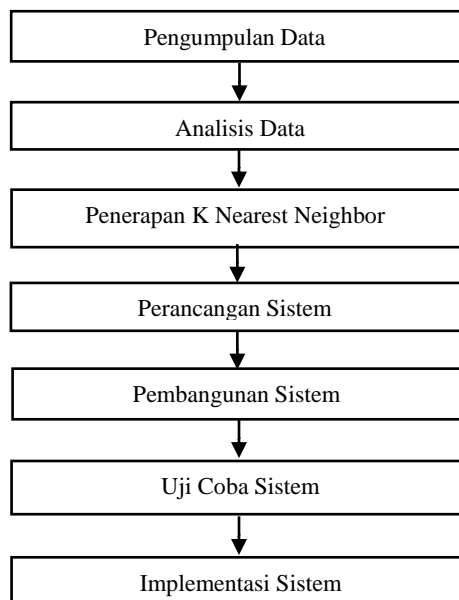
Bersumber pada penelitian sebelumnya yang telah menggunakan salah satu metode *K Nearest Neighbor* yang ditulis oleh Yulia Rizki Amalia[5] yaitu Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode *K Nearest Neighbor* dapat digunakan untuk klasifikasi data penjualan elektronik berdasarkan jumlah penjualannya, yang paling laris. Metode klasifikasi ini dapat membantu pemilik Toko dalam mengelompokkan data.

Hasil dari analisis diatas, maka penulis melakukan penelitian dengan judul “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode *K-Nearest Neighbor*”. Dengan Metode *K-Nearest Neighbor* dapat mempermudah dan membantu pihak UD Andar dalam menentukan produk apa saja yang harus disediakan dengan jumlah yang banyak.

2. METODOLOGI PENELITIAN

2.1 Tahapan Metodologi Penelitian

Tahapan dalam penelitian ini dapat dilihat pada gambar 1 berikut :



Gambar 1. Kerangka Kerja Penelitian

Kerangka kerja yang telah digambarkan diatas, dapat dijabarkan pembahasan masing-masing dalam penelitian adalah sebagai berikut:

1. Pengumpulan Data, Pengumpulan data dilakukan guna memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian.
2. Analisis Data Pada tahap ini dilakukan identifikasi masalah pada sistem yang sedang berjalan. Dengan demikian, diharapkan penulis dapat menemukan kendala-kendala dan permasalahan yang terjadi, sehingga penulis dapat mencari solusi dari permasalahan tersebut.
3. Penerapan *K Nearest Neighbor*, penerapan *k nearest neighbor* merupakan sebuah proses dari data asli (*real*) menjadi data yang diolah menjadi proses perhitungan Algoritma *K Nearest Neighbor*.
4. Perancangan Sistem adalah sebuah kegiatan merancang dan menentukan cara mengolah sistem informasi dari hasil analisa sistem sehingga dapat memenuhi kebutuhan dari pengguna termasuk diantaranya perancangan *user interface*, data dan aktivitas-aktivitas proses. Adapun proses perancangan suatu sistem informasi dilakukan dengan mengamati aliran sistem informasi yang sedang berjalan serta mengubah suatu sistem informasi baru yang telah didesain dengan alat bantu perancangan yaitu *Unified Modelling Language* (UML).
5. Pembangunan Sistem, Membangun sistem informasi merupakan bentuk perubahan organisasi yang direncanakan. Pembangunan sistem merupakan tindakan mengubah, menggantikan atau menyusun sistem secara baik dengan sebagian maupun keseluruhan untuk memperbaiki sistem yang selama ini berjalan.

6. Uji Coba Sistem, Uji coba sistem yaitu suatu proses yang dilakukan untuk menilai apakah sistem yang dibuat telah sesuai dengan apa yang diharapkan. Uji coba sistem pada penelitian ini menggunakan Black Box Testing yaitu metode pengujian sistem perangkat lunak yang tes fungsionalitas dari aplikasi tersebut.
7. Implementasi Sistem, Implementasi sistem adalah prosedur yang dilakukan untuk menyelesaikan desain yang ada dalam dokumen yaitu desain sistem yang disetujui dan menguji, menginstal, memulai, serta menggunakan sistem yang baru atau sistem yang diperbaiki.

2.2 Analisis Data

Analisa data adalah suatu metode pengolahan data menjadi informasi untuk membantu memahami karakteristik data dan memecahkan masalah, terutama yang berkaitan dengan penelitian.

Table 1. Data Penjualan Produk

No	Nama Produk	Kuantitas Produk	Kuantitas Terjual	Kategori
1	Jamu Sidomuncul (kotak)	950	745	Tidak Laris
2	Jamu Klanceng (botol)	600	525	Laris
3	Jamu Asam Urat(kotak)	710	545	Tidak Laris
4	Kantong Plastik Merk Kilo(pk)	915	740	Tidak Laris
5	Kantong Plastik Merk Kristal(pak)	920	850	Laris
6	Kantong Plastik Kamsia(pak)	915	745	Tidak Laris
7	Plastik HD(bks)	1230	1080	Tidak Laris
8	Plastik PE(bks)	1255	1105	Tidak Laris
9	Plastik PP(bks)	1160	1090	Laris
10	Tepung Segitia Biru(kg)	1015	940	Laris
11	Gula pasir(kg)	1245	1170	Laris
12	Mentega (kg)	680	530	Tidak Laris
13	Seres/(kg)	475	400	Laris
14	Pop Ice(pcs)	495	420	Laris
15	Marimas (pcs)	685	610	Laris
16	Best Bubble (kotak)	905	830	Laris
17	Boba(pcs)	530	380	Tidak Laris
18	Susu UHT(btl)	765	690	Laris
19	Susu Kaleng (kaleng)	1115	1045	Laris
20	Sosis Okey (kotak)	1350	1200	Tidak Laris
21	Sosis Asimo (kotak)	1225	1150	Laris
22	Nugget Okey (kotak)	1310	1235	Laris
23	Kulit Kebab(pcs)	1050	900	Tidak Laris
24	Kulit Lumpia (pcs)	845	770	Laris
25	Kentang Goreng Fiesta 1 Kg(kotak)	985	835	Tidak Laris

Table 1 diatas merupakan data penjualan produk yaitu November 2020 - Januari 2022. Data penjualan produk ini didapat dari hasil wawancara kepada pemilik UD Andar.

2.3 Data Mining

Data mining menurut Suyanto adalah gabungan sejumlah disiplin ilmu komputer yang mendefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *database systems*[3]. Berdasarkan Pengertian *Data Mining* menurut para ahli yang disebutkan maka dapat disimpulkan bahwa *Data Mining* adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari *database* yang besar. *Data Mining* merupakan metode untuk menemukan informasi tersembunyi dalam *database* dan bagian dari proses *Knowledge Discovery in Databases* (KDD) untuk menemukan informasi dan pola yang berguna dalam data[4].

Secara umum, Data Mining dibagi menjadi dua kategori utama yaitu[6]:

1. Prediktif
Proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk descriptive mining adalah clustering, asosiasi, dan sequential mining.
2. Deskriptif
Proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan. Salah satu teknik yang terdapat dalam predictive mining adalah klasifikasi. Secara sederhana data mining biasa dikatakan sebagai proses penyaring atau “menambang” pengetahuan dari sejumlah data yang besar.
Secara keseluruhan, proses *Knowledge Discovery in Databases* (KDD) dapat digambarkan sebagai berikut[7]:
1. *Data Selection* (seleksi data)
Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*.

2. *Pre-processing/cleaning* (pembersihan data)
Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.
3. *Transformation* (pengubah data)
Transformation adalah mengubah data kedalam bentuk yang sesuai untuk ditambang.
4. *Data Mining* (Penambangan Data)
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.
5. *Intepretasi/Evaluasi*
Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Seperti menggunakan visualisai atau tempilan yang dapat menjelaskan luaran sistem.

2.4 Penjualan

Penjualan merupakan kegiatan ekonomi normal di mana perusahaan menerima hasil/manfaat yang direncanakan dari penjualan atau pengembalian biaya yang dikeluarkan[4]. Penjualan produk perusahaan ditampilkan setelah dikurangi diskon dan retur[8]. Penjualan adalah pembelian sesuatu (barang atau jasa) dari satu pihak ke pihak lain dengan imbalan uang dari satu pihak[8]. Berdasarkan pengertian penjualan tersebut dapat disimpulkan bahwa penjualan adalah suatu kegiatan ekonomi yang dilakukan antara pembeli dengan penjual melakukan transaksi berupa pertukaran barang atau jasa dengan uang tunai.

2.5 Prediksi

Prediksi/*forecasting* Merupakan salah satu metode yang digunakan untuk memprediksi atau meramalkan suatu hal yang belum pernah terjadi[9]. Prediksi merupakan suatu kegiatan meramalkan secara terstruktur mengenai sesuatu yang mungkin terjadi dimasa akan datang yang bersumber pada informasi dari masa lalu dan sekarang yang dimiliki, agar kesalahannya dapat diperkecil[10]. Berdasarkan Pengertian prediksi tersebut dapat disimpulkan bahwa Prediksi adalah suatu kegiatan peramalan atau perkiraan suatu keadaan dimasa mendatang untuk melakukan sebuah pengujian yang diambil dari data masa lalu dengan tujuan meminimalisir kesalahan yang terjadi.

2.6 Metode K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari label class pada K-NN[4]. Algoritma K-Nearest Neighbor (KNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya[3]. KNN termasuk algoritma *supervised learning*, yang mana hasil dari *query instance* baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi[3].

Tujuan dari algoritma K-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data[4]. Sehingga dengan penerapan algoritma *K-Nearest Neighbor* dapat mempermudah UD Andar pada penjualan produk dengan mengambil objek baru berdasarkan data yang letaknya terdekat dari data baru tersebut. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidian[11].

Algoritma K-NN terdapat 5 (lima) cara, untuk mencari tetangga terdekat, yaitu: [3]

1. Jarak Euclidean
2. Jarak Manhattan
3. Jarak Cosine
4. Jarak Correlation
5. Jarak Hamming

Penelitian ini penulis hanya menggunakan jarak Euclidean, maka rumus perhitungan jarak dengan Euclidean berikut ini, Sayad[3].

$$\sqrt{\sum_{i=1}^K (X_i - Y_i)^2 + (X_i - Y_i)^2 + \dots + \dots} = \dots \dots \dots (1)$$

Keterangan Rumus:

Nilai X_i : nilai yang ada pada data *training*

Nilai Y_i : nilai yang ada pada data *testing* .

Nilai K merupakan dimensi atribut.

Langkah-langkah perhitungan *K Nearest Neighbor*(KNN) sebagai berikut:

1. Menentukan parameter K (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak Eucliden objek terhadap data training yang diberikan
3. Selanjutnya mengurutkan hasil no 2 secara *ascending* (berurutan dari nilai tinggi ke rendah)
4. Mengumpulkan kategori Y (klasifikasi nearest neighbor berdasarkan nilai k)

5. Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksi objek yang baru. (k) pada algoritma *k-nearest neighbor* adalah banyaknya tetangga terdekat yang akan digunakan sebagai titik untuk melakukan klasifikasi pada data atau objek baru. Untuk menghitung jarak antar objek data pada algoritma *k-nearest neighbor* dapat dilakukan dengan beberapa cara, salah satunya adalah dengan *Ecludiean Distance*.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan K-Nearest Neighbor Untuk Prediksi Penjualan Produk Terlaris

Berdasarkan ketentuan dalam penentuan produk terlaris maka didapatkan hasil transformasi data. Berikut data yang ditampilkan dalam bentuk table 2 berikut:

Tabel 2. Transformasi Data

No	Nama Produk	Kuantitas Produk	Kuantitas Terjual	Kategori	Satuan
1	Jamu Sidomuncul	950	745	Tidak Laris	Kotak
2	Jamu Klanceng	600	525	Laris	Botol
3	Jamu Asam Urat	710	545	Tidak Laris	Kotak
4	Kantong Plastik Merk Kilo	915	740	Tidak Laris	Pak
5	Kantong Plastik Merk Kristal	920	850	Laris	Pak
6	Kantong Plastik Kamsia	915	745	Tidak Laris	Pak
7	Plastik HD	1230	1080	Tidak Laris	Bungkus
8	Plastik PE	1255	1105	Tidak Laris	Bungkus
9	Plastik PP	1160	1090	Laris	Bungkus
10	Tepung Segitia Biru	1015	940	Laris	Kilogram
11	Gula pasir	1245	1170	Laris	Kilogram
12	Mentega	680	530	Tidak Laris	Kilogram
13	Seres	475	400	Laris	Kilogram
14	Pop Ice	495	420	Laris	Pcs
15	Marimas	685	610	Laris	Pcs
16	Best Bubble	905	830	Laris	Kotak
17	Boba	530	380	Tidak Laris	Pcs
18	Susu UHT	765	690	Laris	Botol
19	Susu Kaleng	1115	1045	Laris	Kaleng
20	Sosis Okey	1350	1200	Tidak Laris	Kotak
21	Sosis Asimo	1225	1150	Laris	Kotak
22	Nugget Okey	1310	1235	Laris	Kotak
23	Kulit Kebab	1050	900	Tidak Laris	Pcs
24	Kulit Lumpia	845	770	Laris	Pcs
25	Kentang Goreng Fiesta 1 Kg	985	835	Tidak Laris	Kotak

Berdasarkan transformasi data tersebut, langkah selanjutnya yaitu melakukan normalisasi data dengan menggunakan normalisasi min max. dengan menggunakan normalisasi min max ini maka hasil akurasi menjadi lebih tepat dibandingkan tanpa menggunakan normalisasi. Berikut rumus pada normalisasi min max yaitu :

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Keterangan :

x' : nilai skala baru

x : nilai data asli sebelum dilakukan normalisasi

min : nilai terendah dari data

max : nilai tertinggi dari data

Berikut penyelesaian perhitungannya yang dibuat dalam bentuk tabel:

Nilai Min untuk Kuantitas Produk :475

Nilai Max Untuk Kuantitas Produk:1350

Nilai Min Untuk Kuantitas Terjual :380

Nilai Min Untuk Kuantitas Terjual :1235

Tabel 3. Table Data Normalisasi Min Max

No	Min Max Kuantitas Produk	Hasil	Min Max Kuantitas Terjual	Hasil
1	$950 - 475 / 1350 - 475 =$	0,5429	$745 - 380 / 1235 - 380 =$	0,4269
2	$600 - 475 / 1350 - 475 =$	0,1429	$525 - 380 / 1235 - 380 =$	0,1696

No	Min Max Kuantitas Produk	Hasil	Min Max Kuantitas Terjual	Hasil
3	710 – 475/1350 – 475 =	0,2686	545 – 380/1235 – 380 =	0,193
4	915 – 475/1350 – 475 =	0,5029	740 – 380/1235 – 380 =	0,4211
5	920 – 475/1350 – 475 =	0,5086	850 – 380/1235 – 380 =	0,5497
6	915 – 475/1350 – 475 =	0,5029	745 – 380/1235 – 380 =	0,4269
7	1230 – 475/1350 – 475 =	0,8629	1080 – 380/1235 – 380 =	0,8187
8	1255 – 475/1350 – 475 =	0,8914	1105 – 380/1235 – 380 =	0,848
9	1160 – 475/1350 – 475 =	0,7829	1090 – 380/1235 – 380 =	0,8304
10	1015 – 475/1350 – 475 =	0,6171	940 – 380/1235 – 380 =	0,655
11	1245 – 475/1350 – 475 =	0,88	1170 – 380/1235 – 380 =	0,924
12	680 – 475/1350 – 475 =	0,2343	530 – 380/1235 – 380 =	0,1754
13	475 – 475/1350 – 475 =	0	400 – 380/1235 – 380 =	0,0234
14	495 – 475/1350 – 475 =	0,0229	420 – 380/1235 – 380 =	0,0468
15	685 – 475/1350 – 475 =	0,24	610 – 380/1235 – 380 =	0,269
16	905 – 475/1350 – 475 =	0,4914	830 – 380/1235 – 380 =	0,5263
17	530 – 475/1350 – 475 =	0,0629	380 – 380/1235 – 380 =	0
18	765 – 475/1350 – 475 =	0,3314	690 – 380/1235 – 380 =	0,3626
19	1115 – 475/1350 – 475 =	0,7314	1045 – 380/1235 – 380 =	0,7778
20	1350 – 475/1350 – 475 =	1	1200 – 380/1235 – 380 =	0,9591
21	1225 – 475/1350 – 475 =	0,8571	1150 – 380/1235 – 380 =	0,9006
22	1310 – 475/1350 – 475 =	0,9543	1235 – 380/1235 – 380 =	1
23	1050 – 475/1350 – 475 =	0,6571	900 – 380/1235 – 380 =	0,6082
24	845 – 475/1350 – 475 =	0,4229	770 – 380/1235 – 380 =	0,4561
25	985 – 475/1350 – 475 =	0,5829	835 – 380/1235 – 380 =	0,5322
26	400 – 475/1350 – 475 =	-0,0857	395 – 380/1235 – 380 =	0,0175

Proses perhitungan *K Nearest Neighbor* yaitu :

1. Menentukan parameter K = jumlah tetangga terdekat. Pada penelitian ini nilai k yang digunakan adalah 7.
2. Hitung jarak antara data testing (uji) dengan semua data training pada tahap transformasi menggunakan perhitungan jarak Euclidean Distance, yang dipakai untuk menghitung jarak Euclidean Distance [7].

Tabel 4. Hasil dari Jarak *Euclidean Distance*

No	Nama Produk	Kuantitas Produk	Kuantitas Terjual	Jarak Euclidean
1	Jamu Sidomuncul (kotak)	950	745	$\sqrt{(0,5429 - (-0,0857))^2 + (0,4269 - 0,0175)^2} = 0,7501$
2	Jamu Klanceng (botol)	600	525	$\sqrt{(0,1429 - (-0,0857))^2 + (0,1696 - 0,0175)^2} = 0,2745$
3	Jamu Asam Urat(kotak)	710	545	$\sqrt{(0,2686 - (-0,0857))^2 + (0,193 - 0,0175)^2} = 0,3953$
4	Kantong Plastik Merk Kilo(pk)	915	740	$\sqrt{(0,5029 - (-0,0857))^2 + (0,4211 - 0,0175)^2} = 0,7136$
5	Kantong Plastik Merk Kristal(pak)	920	850	$\sqrt{(0,5086 - (-0,0857))^2 + (0,5497 - 0,0175)^2} = 0,7977$
6	Kantong Plastik Kamsia(pak)	915	745	$\sqrt{(0,5029 - (-0,0857))^2 + 4269} = 0,7169$
7	Plastik HD(bks)	1230	1080	$\sqrt{(0,8629 - (-0,0857))^2 + (0,8187 - 0,0175)^2} = 1,2416$
8	Plastik PE(bks)	1255	1105	$\sqrt{(0,8914 - (-0,0857))^2 + (0,848 - 0,0175)^2} = 1,2823$
9	Plastik PP(bks)	1160	1090	$\sqrt{(0,7829 - (-0,0857))^2 + (0,8304 - 0,0175)^2} = 1,1896$
10	Tepung Segitia Biru(kg)	1015	940	$\sqrt{(0,6171 - (-0,0857))^2 + (0,655 - 0,0175)^2} = 0,9489$
11	Gula pasir(kg)	1245	1170	$\sqrt{(0,88 - (-0,0857))^2 + 924^2} = 1,3245$
12	Mentega (kg)	680	530	$\sqrt{(0,2343 - (-0,0857))^2 + (0,1754 - 0,0175)^2} = 0,3568$
13	Seres/(kg)	475	400	$\sqrt{(0 - (-0,0857))^2 + (0,0234 - 0,0175)^2} = 0,0859$
14	Pop Ice(pcs)	495	420	$\sqrt{(0,0229 - (-0,0857))^2 + (0,0468 - 0,0175)^2} = 0,1124$
15	Marimas (pcs)	685	610	$\sqrt{(0,24 - (-0,0857))^2 + (0,269 - 0,0175)^2} = 0,4115$
16	Best Bubble (kotak)	905	830	$\sqrt{(0,4914 - (-0,0857))^2 + (0,5263 - 0,0175)^2} = 0,7694$
17	Boba(pcs)	530	380	$\sqrt{(0,0629 - (-0,0857))^2 + (0 - 0,0175)^2} = 0,1496$
18	Susu UHT(btl)	765	690	$\sqrt{(0,3314 - (-0,0857))^2 + (0,3626 - 0,0175)^2} = 0,5413$
19	Susu Kaleng (kaleng)	1115	1045	$\sqrt{(0,7314 - (-0,0857))^2 + (0,7778 - 0,0175)^2} = 1,1161$
20	Sosis Okey (kotak)	1350	1200	$\sqrt{(1 - (-0,0857))^2 + (0,9591 - 0,0175)^2} = 1,4371$
21	Sosis Asimo (kotak)	1225	1150	$\sqrt{(0,8751 - (-0,0857))^2 + (0,9006 - 0,0175)^2} = 1,2918$
22	Nugget Okey (kotak)	1310	1235	$\sqrt{(0,9543 - (-0,0857))^2 + (1 - 0,0175)^2} = 1,4307$
23	Kulit Kebab(pcs)	1050	900	$\sqrt{(0,6571 - (-0,0857))^2 + (0,6082 - 0,0175)^2} = 0,9491$
24	Kulit Lumpia (pcs)	845	770	$\sqrt{(0,4229 - (-0,0857))^2 + (0,4561 - 0,0175)^2} = 0,6716$
25	Kentang Goreng Fiesta 1 Kg(kotak)	985	835	$\sqrt{(0,5829 - (-0,0857))^2 + (0,5322 - 0,0175)^2} = 0,8437$

3. Mengurutkan hasil secara *ascending* (dari nilai tinggi ke rendah)Tabel 5. Urutan *ascending*

No	Jarak Euclidean	Rangking
1	$\sqrt{(0,5429 - (-0,0857))^2 + (0,4269 - 0,0175)^2}$ = 0,7501	12
2	$\sqrt{(0,1429 - (-0,0857))^2 + (0,1696 - 0,0175)^2}$ = 0,2745	4
3	$\sqrt{(0,2686 - (-0,0857))^2 + (0,193 - 0,0175)^2}$ = 0,3953	6
4	$\sqrt{(0,5029 - (-0,0857))^2 + (0,4211 - 0,0175)^2}$ = 0,7136	10
5	$\sqrt{(0,5086 - (-0,0857))^2 + (0,5497 - 0,0175)^2}$ = 0,7977	14
6	$\sqrt{(0,5029 - (-0,0857))^2 + 4269}$ = 0,7169	11
7	$\sqrt{(0,8629 - (-0,0857))^2 + (0,8187 - 0,0175)^2}$ = 1,2416	20
8	$\sqrt{(0,8914 - (-0,0857))^2 + (0,848 - 0,0175)^2}$ = 1,2823	21
9	$\sqrt{(0,7829 - (-0,0857))^2 + (0,8304 - 0,0175)^2}$ = 1,1896	19
10	$\sqrt{(0,6171 - (-0,0857))^2 + (0,655 - 0,0175)^2}$ = 0,9489	16
11	$\sqrt{(0,88 - (-0,0857))^2 + 924^2}$ = 1,3245	23
12	$\sqrt{(0,2343 - (-0,0857))^2 + (0,1754 - 0,0175)^2}$ = 0,3568	5
13	$\sqrt{(0 - (-0,0857))^2 + (0,0234 - 0,0175)^2}$ = 0,0859	1
14	$\sqrt{(0,0229 - (-0,0857))^2 + (0,0468 - 0,0175)^2}$ = 0,1124	2
15	$\sqrt{(0,24 - (-0,0857))^2 + (0,269 - 0,0175)^2}$ = 0,4115	7
16	$\sqrt{(0,4914 - (-0,0857))^2 + (0,5263 - 0,0175)^2}$ = 0,7694	13
17	$\sqrt{(0,0629 - (-0,0857))^2 + (0 - 0,0175)^2}$ = 0,1496	3
18	$\sqrt{(0,3314 - (-0,0857))^2 + (0,3626 - 0,0175)^2}$ = 0,5413	8
19	$\sqrt{(0,7314 - (-0,0857))^2 + (0,7778 - 0,0175)^2}$ = 1,1161	18
20	$\sqrt{(1 - (-0,0857))^2 + (0,9591 - 0,0175)^2}$ = 1,4371	25
21	$\sqrt{(0,8751 - (-0,0857))^2 + (0,9006 - 0,0175)^2}$ = 1,2918	22
22	$\sqrt{(0,9543 - (-0,0857))^2 + (1 - 0,0175)^2}$ = 1,4307	24
23	$\sqrt{(0,6571 - (-0,0857))^2 + (0,6082 - 0,0175)^2}$ = 0,9491	17
24	$\sqrt{(0,4229 - (-0,0857))^2 + (0,4561 - 0,0175)^2}$ = 0,6716	9
25	$\sqrt{(0,5829 - (-0,0857))^2 + (0,5322 - 0,0175)^2}$ = 0,8437	15
26	0	

Selanjutnya, menentukan jarak *Euclidean* untuk mendapatkan nilai prediksi data uji dan tetangga terdekatnya. Berikut perhitungan jarak *Euclidean* yang dibuat dalam bentuk table. Simbol L (Terlaris), dan TL (Tidak Laris) :

Tabel 6. Prediksi *K Nearest Neighbor*

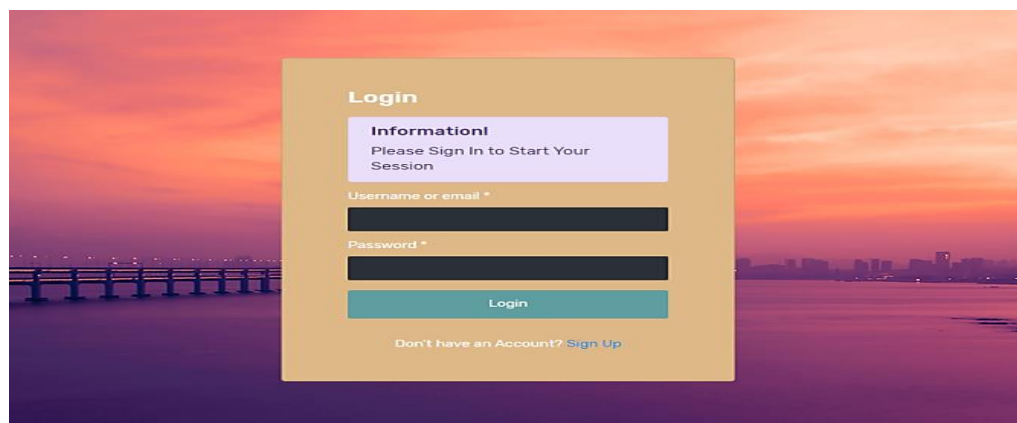
No	Jarak <i>Euclidean</i>	Rank	K=1	K=3	K=5	K=7
1	$\sqrt{(0,5429 - (-0,0857))^2 + (0,4269 - 0,0175)^2} = 0,7501$	12				
2	$\sqrt{(0,1429 - (-0,0857))^2 + (0,1696 - 0,0175)^2} = 0,2745$	4			L	L
3	$\sqrt{(0,2686 - (-0,0857))^2 + (0,193 - 0,0175)^2} = 0,3953$	6				TL
4	$\sqrt{(0,5029 - (-0,0857))^2 + (0,4211 - 0,0175)^2} = 0,7136$	10				
5	$\sqrt{(0,5086 - (-0,0857))^2 + (0,5497 - 0,0175)^2} = 0,7977$	14				
6	$\sqrt{(0,5029 - (-0,0857))^2 + 4269} = 0,7169$	11				
7	$\sqrt{(0,8629 - (-0,0857))^2 + (0,8187 - 0,0175)^2} = 1,2416$	20				
8	$\sqrt{(0,8914 - (-0,0857))^2 + (0,848 - 0,0175)^2} = 1,2823$	21				
9	$\sqrt{(0,7829 - (-0,0857))^2 + (0,8304 - 0,0175)^2} = 1,1896$	19				
10	$\sqrt{(0,6171 - (-0,0857))^2 + (0,655 - 0,0175)^2} = 0,9489$	16				
11	$\sqrt{(0,88 - (-0,0857))^2 + 924^2} = 1,3245$	23				
12	$\sqrt{(0,2343 - (-0,0857))^2 + (0,1754 - 0,0175)^2} = 0,3568$	5			TL	TL
13	$\sqrt{(0 - (-0,0857))^2 + (0,0234 - 0,0175)^2} = 0,0859$	1	L	L	L	L
14	$\sqrt{(0,0229 - (-0,0857))^2 + (0,0468 - 0,0175)^2} = 0,1124$	2		L	L	L
15	$\sqrt{(0,24 - (-0,0857))^2 + (0,269 - 0,0175)^2} = 0,4115$	7				L
16	$\sqrt{(0,4914 - (-0,0857))^2 + (0,5263 - 0,0175)^2} = 0,7694$	13				
17	$\sqrt{(0,0629 - (-0,0857))^2 + (0 - 0,0175)^2} = 0,1496$	3		TL	TL	TL
18	$\sqrt{(0,3314 - (-0,0857))^2 + (0,3626 - 0,0175)^2} = 0,5413$	8				
19	$\sqrt{(0,7314 - (-0,0857))^2 + (0,7778 - 0,0175)^2} = 1,1161$	18				
20	$\sqrt{(1 - (-0,0857))^2 + (0,9591 - 0,0175)^2} = 1,4371$	25				
21	$\sqrt{(0,8751 - (-0,0857))^2 + (0,9006 - 0,0175)^2} = 1,2918$	22				
22	$\sqrt{(0,9543 - (-0,0857))^2 + (1 - 0,0175)^2} = 1,4307$	24				
23	$\sqrt{(0,6571 - (-0,0857))^2 + (0,6082 - 0,0175)^2} = 0,9491$	17				
24	$\sqrt{(0,4229 - (-0,0857))^2 + (0,4561 - 0,0175)^2} = 0,6716$	9				
25	$\sqrt{(0,5829 - (-0,0857))^2 + (0,5322 - 0,0175)^2} = 0,8437$	15				
26	0		L	L	L	L

Berdasarkan hasil perhitungan jarak diatas, maka didapatkan suatu hasil dengan nilai k=7 memiliki keputusan yaitu Terlaris = 4 dan Tidak Laris = 3. Dapat dilihat pada Tabel 4.6 mayoritas klasifikasi yang memiliki jumlah paling banyak adalah kategori “Terlaris”. Sehingga penjualan produk memiliki nilai sesuai dengan data uji yang telah dihitung diprediksi masuk kedalam kategori “Terlaris” adalah “Produk Tepung segitia Biru”.

3.2 Implementasi Hasil

Perhitungan prediksi penjualan dapat dilakukan menggunakan aplikasi sistem yang sudah dibuat kedalam website. Peneliti menggunakan data training yang sekaligus menjadi data testing. Berikut tampilan aplikasi perhitungan prediksi penjualan.

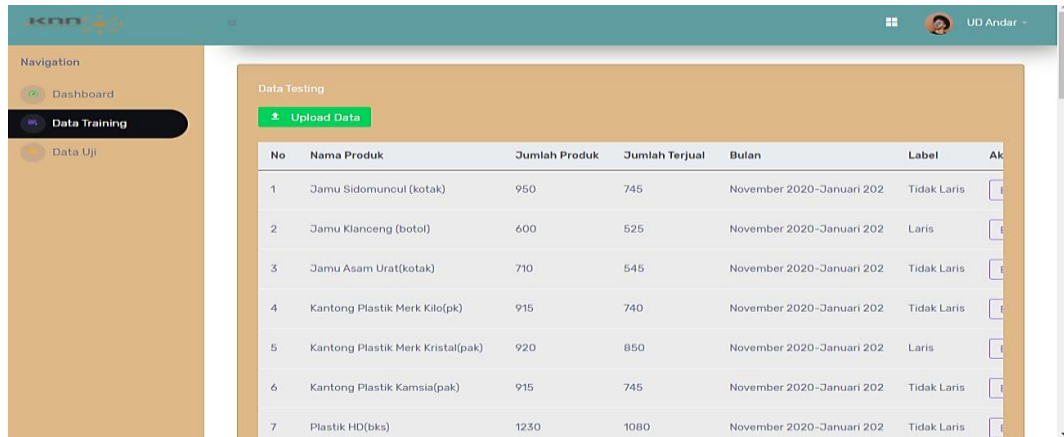
1. Tampilan Login



Gambar 2. Tampilan Login

Gambar 2 diatas merupakan Tampilan halaman login dimana untuk memperjelas hak administrator masuk ke pemrosesan sistem. Administrator dengan username dan password yang valid berhak masuk ke sistem melalui tampilan halaman login.

2. Tampilan Menu Data *Training*

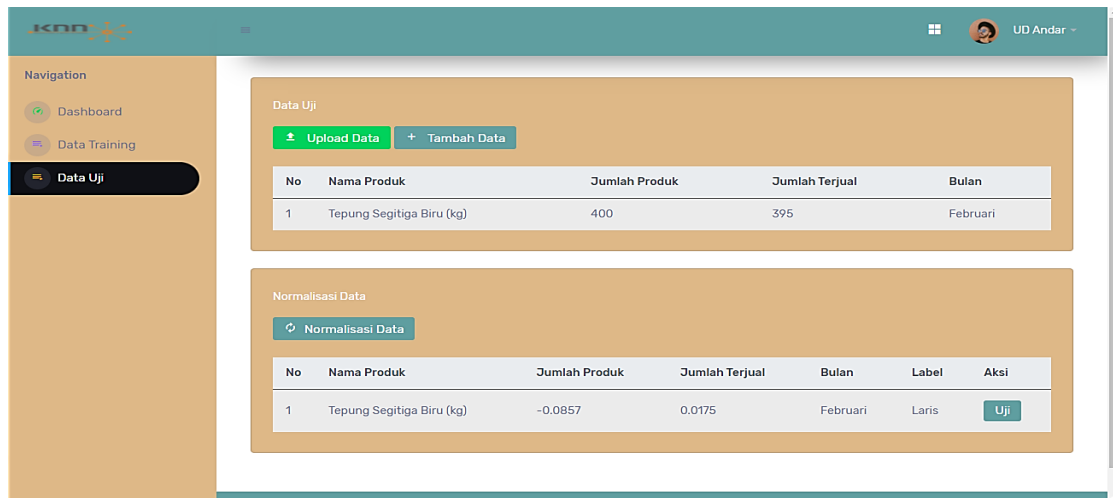


No	Nama Produk	Jumlah Produk	Jumlah Terjual	Bulan	Label	Aksi
1	Jamu Sidomuncul (kotak)	950	745	November 2020-Januari 202	Tidak Laris	
2	Jamu Klanceng (botol)	600	525	November 2020-Januari 202	Laris	
3	Jamu Asam Urat(kotak)	710	545	November 2020-Januari 202	Tidak Laris	
4	Kantong Plastik Merk Kilo(pk)	915	740	November 2020-Januari 202	Tidak Laris	
5	Kantong Plastik Merk Kristal(pak)	920	850	November 2020-Januari 202	Laris	
6	Kantong Plastik Kamsia(pak)	915	745	November 2020-Januari 202	Tidak Laris	
7	Plastik HD(bks)	1230	1080	November 2020-Januari 202	Tidak Laris	

Gambar 3. Tampilan Menu Data *Training*

Gambar 3 merupakan Tampilan menu data testing dimana halaman ini memuat data penjualan produk yang diambil dari November 2020 hingga Januari 2022. Data yang didapat dari hasil wawancara tersebut dicatat melalui microsoft excel, kemudian untuk import data tersebut cukup mengupload data, maka data yang dari excel sebelumnya sudah tersimpan didalam database dan ditampilkan disistem.

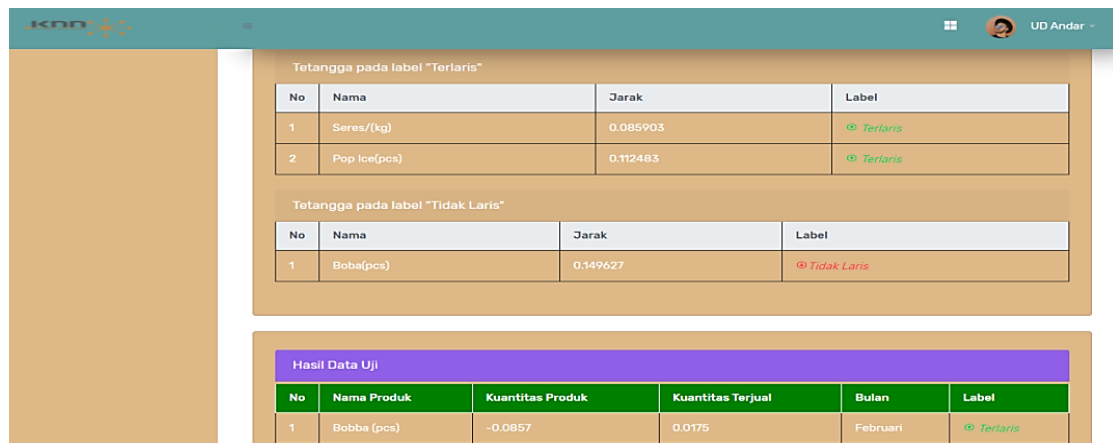
3. Tampilan Menu Data Uji



No	Nama Produk	Jumlah Produk	Jumlah Terjual	Bulan
1	Tepung Segitiga Biru (kg)	400	395	Februari

No	Nama Produk	Jumlah Produk	Jumlah Terjual	Bulan	Label	Aksi
1	Tepung Segitiga Biru (kg)	-0.0857	0.0175	Februari	Laris	Uji

Gambar 4. Tampilan Menu Data Uji



No	Nama	Jarak	Label
1	Seres/(kg)	0.085903	Terlaris
2	Pop Ice(pcs)	0.112483	Terlaris

No	Nama	Jarak	Label
1	Boba(pcs)	0.149627	Tidak Laris

No	Nama Produk	Kuantitas Produk	Kuantitas Terjual	Bulan	Label
1	Boba (pcs)	-0.0857	0.0175	Februari	Terlaris

Gambar 5. Tampilan Hasil Perhitungan

Gambar 4 dan 5 merupakan Tampilan menu pada data uji dimana halaman ini dapat menambah data produk dengan menginputkan satu persatu atau juga bisa dengan mengupload data dari excel. Pada halaman ini akan dilakukan pengujian dengan proses perhitungan *k nearest neighbor*, kemudian hasil dari perhitungan tersebut ditampilkan pada halaman ini.

4. KESIMPULAN

Berdasarkan penelitian ini dapat disimpulkan bahwa perhitungan dengan teknik data mining dan algoritma *k nearest neighbor* didapatkan hasil prediksi dengan nilai akurasi yang tinggi. Dengan menerapkan metode *k nearest neighbor* kedalam sebuah sistem aplikasi maka dapat membantu UD tersebut dalam menyelesaikan masalah yang dihadapi, sehingga teknik data mining dan metode algoritma *k nearest neighbor* ini dapat diimplementasikan untuk memprediksi penjualan produk terlaris pada UD Andar.

REFERENCES

- [1] A. Alfani W.P.R., F. Rozi, and F. Sukmana, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 1, pp. 155–160, 2021, doi: 10.29100/jipi.v6i1.1910.
- [2] 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.
- [3] S. Rahmatullah, S. Wahyuni, M. F. Chaining, and F. C. Method, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Furniture Terlaris Menggunakan Metode Knearest Neighbor," no. 2, pp. 75–86, 2020.
- [4] Hasmawati, J. Nangi, and M. Muchtar, "Aplikasi prediksi penjualan barang menggunakan metode k- nearest neighbor (knn) (studi kasus tumaka mart)," *semanTIK*, vol. 3, no. Jul-Des 2017, pp. 151–160, 2019.
- [5] Y. R. Amalia, "Penerapan data mining untuk prediksi penjualan produk elektronik terlaris menggunakan metode k-nearest neighbor," 2018.
- [6] N. Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.
- [7] W. M. S. Karsito, "PREDIKSI POTENSI PENJUALAN PRODUK DELIFRANCE DENGAN METODE NAIVE BAYES DI PT. PANGAN LESTARI," vol. 9, no. September, pp. 67–78, 2018.
- [8] Andi Saputra, Ashari Imamuddin, and Pria Sukamto, "Rancang Bangun Aplikasi Sistem Penjualan Case Study: Pt. X," *INFOTECH J. Inform. Teknol.*, vol. 1, no. 2, pp. 78–86, 2020, doi: 10.37373/infotech.v1i2.67.
- [9] I. Firnando, "Implementasi Algoritma Apriori Dan Forecasting," *Pros. SENTIA*, vol. 7, no. 3, pp. 2085–2347, 2019.
- [10] J. Informasi and D. Komputer, "PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN PRODUK TRIPLEK PADA PT PUNCAK MENARA HIJAU MAS," *J. Inf. dan Komput.*, vol. 8, 2020.
- [11] Y. Yahya and W. Puspita Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020, doi: 10.29408/jit.v3i2.2279.