JOISIE (Journal Of Information Systems And Informatics Engineering)
Vol. 8, No.2, Desember 2024, Hlm 259-271
Received: 29 Oktober 2024, Revised: 29 November 2024, Accepted:9 Desember 2024
p- ISSN: 2503-5304
e- ISSN: 2527-3116

PENERAPAN *DATA MINING* UNTUK PREDIKSI PENJUALAN OBAT MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

Syabilla Mutia Rahma[1)], Nabila Rizky Oktadini[2*)], Ken Ditha Tania[3)], Dwi Rosa Indah[4)]
[1,2,3,4]Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang
email: [1]syabillamutiara26@gmail.com, [2]nabilarizky@unsri.ac.id, [3]kenya.tania@gmail.com, [4]indah812@unsri.ac.id

Abstract

PT XYZ Palembang is one of the national pharmaceutical companies that supplies medicine in Indonesia. This company produces various medicinal products in very large quantities. It is known that sales of medicinal products in the last 3 years have decreased and increased every month. With the decrease and increase in sales, the company must estimate drug sales well to ensure that the availability of drug stock is not excessive or lacking. The researcher proposed the implementation of the K-Nearest Neighbor algorithm with the aim of determining the results of predicting the number of drug sales at PT XYZ Palembang. K-Nearest Neighbor is applied through manual calculations and using RapidMiner software. The prediction results obtained through RapidMiner produce a good level of accuracy, reaching 90%. With these prediction results, it is revealed that the K-Nearest Neighbor method is effective in predicting future sales. The K-NN model can be one solution to predict drug sales based on previously existing sales data.

Keywords: *Sales Prediction, K-Nearest Neighbor, Data Mining , RapidMiner*

Abstrak

PT XYZ Palembang merupakan salah satu perusahaan farmasi nasional yang menyuplai pemenuhan obat di Indonesia. Perusahaan ini memproduksi berbagai produk obat-obatan dengan jumlah yang sangat besar. Diketahui bahwa penjualan produk obat-obatan dalam 3 tahun terakhir mengalami penurunan dan peningkatan setiap bulan. Dengan adanya penurunan dan peningkatan penjualan tersebut, perusahaan harus memperkirakan penjualan obat dengan baik untuk memastikan ketersediaan stok obat tidak berlebihan maupun kekurangan. Peneliti mengusulkan implementasi algoritma *K-Nearest Neighbor* dengan tujuan untuk mengetahui hasil prediksi jumlah penjualan obat-obatan pada PT XYZ Palembang. *K-Nearest Neighbor* diterapkan melalui perhitungan manual serta dengan menggunakan *software RapidMiner*. Hasil prediksi yang diperoleh melalui *RapidMiner* menghasilkan tingkat akurasi baik yaitu mencapai 90%. Dengan hasil prediksi tersebut mengungkapkan bahwa metode *K-Nearest Neighbor* efektif diterapkan untuk melakukan prediksi penjualan di masa mendatang. Model K-NN dapat menjadi salah satu solusi untuk memprediksi penjualan obat-obatan berdasarkan data penjualan yang telah ada sebelumnya.

Kata Kunci: Prediksi Penjualan, *K-Nearest Neighbor*, *Data Mining*, *RapidMiner*

1. PENDAHULUAN

Ketepatan dalam memastikan ketersediaan produk merupakan aspek penting dalam keberhasilan pada ranah operasi penjualan. Dalam bidang kesehatan, pentingnya mempertahankan tingkat pasokan obat yang optimal tidak dapat disangkal karena pasokan obat yang memadai dapat memenuhi kebutuhan dalam penyedia layanan kesehatan. Hal ini menggarisbawahi pentingnya presisi dan efisiensi dalam mengelola ketersediaan obat untuk menjadikan fungsi proses penjualan dan pembelian yang baik di sektor kesehatan. Dari kasus tersebut menjadi salah satu faktor pendukung dalam penjualan obat di seluruh perusahaan farmasi di Indonesia, salah satunya PT XYZ Palembang yang merupakan perusahaan farmasi nasional dalam suplai pemenuhan obat di Indonesia.

Persaingan dalam dunia bisnis mengharuskan para pengusaha untuk menciptakan strategi yang dapat meningkatkan penjualan produk mereka (Rismala et al., 2023). Salah satu strategi yang efektif untuk mempersiapkan diri menghadapi situasi mendatang adalah dengan membuat prediksi yang

https://doi.org/10.35145/joisie.v8i2.4725

akurat tentang kondisi tersebut. Prediksi ini akan membantu memperkirakan jumlah penjualan obat yang paling mungkin terjadi di masa depan (Putri and Andri, 2022). Mengetahui nilai penjualan di masa depan akan membantu perusahaan untuk memahami jenis obat apa yang memiliki kebutuhan lebih dan mana yang tidak.

Prediksi merupakan proses meramalkan secara sistematis kemungkinan kejadian masa depan, didasarkan pada pengetahuan yang diperoleh dari data sebelumnya dan saat ini, dengan tujuan utuk mengurangi perbedaan antara keadaan aktual dan hasil yang di prediksikan. Prediksi tidak selalu menghasilkan solusi yang tepat mengenai kejadian di masa depan, tetapi berupaya untuk mengidentifikasi perkiraan yang lebih dekat dengan apa yang mungkin terjadi (Pohan et al., 22AD). Prediksi penjualan merupakan komponen penting dalam strategi bisnis. karena menetapkan kerangka anggaran penjualan yang berpengaruh terhadap inovasi produk, perencanaan pengeluaran operasional, perkiraan laba rugi, serta keseimbangan keuangan perusahaan (Duran et al., 2024).

Salah satu manfaat dari melakukan prediksi adalah memberikan panduan bagi perusahaan dalam memutuskan berapa banyak jumlah produk yang perlu disediakan. Prediksi juga berperan penting dalam merencanakan persediaan, karena dapat menghasilkan informasi yang diharapkan dapat mengurangi kemungkinan terjadinya kesalahan dalam perencanaan. Prediksi umumnya dimanfaatkan untuk mengambil informasi dari sekumpulan data dalam jumlah besar, sehingga diperlukan adanya penerapan teknik *data mining*.

PT XYZ Palembang merupakan perusahaan farmasi nasional yang menyuplai pemenuhan obat untuk memprioritaskan kebutuhan kesehatan masyarakat Indonesia. Diketahui bahwa penjualan produk obat-obatan ini mengalami penurunan dan peningkatan setiap bulan. Dengan adanya fluktuasi penjualan tersebut, perusahaan harus merencanakan dan menyiapkan penjualan obat di periode berikutnya. Permasalahan sering muncul dalam memperkirakan penjualan obat untuk periode berikutnya. Kesalahan dalam memperkirakan penjualan obat dapat mengakibatkan kelebihan persediaan obat, sehingga berpotensi menyebabkan kualitas produk menurun karena kedaluwarsa atau kurangnya ketersediaan stok yang dapat menurunkan laba perusahaan.

Maka dari itu, penelitian ini mengajukan solusi untuk dapat melakukan prediksi penjualan obat periode berikutnya melalui penerapan *data mining*. *Data mining* merupakan suatu tahapan yang mencakup eksplorasi dan analisis terhadap data guna menemukan pola-pola tersembunyi atau informasi penting yang bisa digunakan untuk menghasilkan keputusan atau tindakan yang lebih tepat. *Data mining* mencakup penerapan berbagai teknik analisis data yang kompleks untuk mengeluarkan informasi dari sekumpulan data yang sangat besar, beragam, dan kompleks (Rismala et al., 2023). Proses pada data mining memanfaatkan beberapa metode pembelajaran mesin (*machine learning*) untuk meneliti dan mengekstraksi data secara otomatis. *Data mining* mampu mengenali pola dan interaksi kumpulan data dalam jumlah besar, yang mengarah pada penemuan informasi atau pengetahuan dalam basis data (Jessfry and Siddik, 2024). Maka, *data mining* merupakan sebuah informasi yang diperlukan untuk mengekstrak data-data penjualan produk obat dan selanjutnya dikelola untuk dapat menghasilkan suatu prediksi pada periode kedepannya. Satu dari beberapa pendekatan yang bisa diterapkan pada *data mining* untuk melakukan prediksi terhadap data adalah metode *K-Nearest Neighbor*.

Metode *K-Nearest Neighbor* (K-NN) merupakan sebuah metode untuk melakukan pengklasifikasian atau pengelompokkan data dengan mempertimbangkan jarak antara satu data dengan data lainnya. Meskipun K-NN termasuk dalam kategori pendekatan yang relatif sederhana, metode ini menunjukkan tingkat akurasi yang baik (Yolanda and Fahmi, 2021). Metode *K-Nearest Neighbor* berupaya untuk mengkategorikan data baru yang kelas nya belum teridentifikasi dengan memilih beberapa data yang paling mendekati data tersebut. Untuk dapat menghitung jarak terdekat, dataset awalnya dipisahkan menjadi data pelatihan (*training*) dan data uji (*testing*). Setelah data pelatihan dan data uji ditetapkan, jarak setiap data *testing* dan data *training* dihitung memanfaatkan teknik *Euclidean Distance*. Algoritma K-NN memiliki tujuan untuk mengelompokkan atau mengategorikan objek baru dengan memanfaatkan atribut dan data pelatihan yang ada (Nangi et al., 2019).

Metode *K-Nearest Neighbor* sudah sering digunakan pada penelitian yang berhubungan dengan prediksi. Seperti penelitian yang dilakukan oleh (Putri, 2021) untuk melakukan prediksi pada penjualan buah dan sayur di PT Central Brastagi Utama. Selain pada bidang penjualan, metode ini juga pernah digunakan untuk prediksi kelulusan tepat waktu berdasarkan riwayat akademik (Riadi et

https://doi.org/10.35145/joisie.v8i2.4725

al., 2024), perkiraan angka kelahiran berdasarkan berbagai kelompok usia ibu (Syafana et al., 2024), prediksi penyakit diabetes pada wanita (Silalahi et al., 2023), klasifikasi *tweets* pada *twitter* (Ramadhona Hassolthine et al., 2023), serta penelitian yang lainnya. Adapun yang menjadi perbedaan dari penelitian sebelumnya yaitu penelitian ini berfokus kepada prediksi penjualan obat pada PT XYZ Palembang menggunakan metode K-NN dengan menerapkan *software RapidMiner*.

Penelitian ini dilakukan dengan tujuan untuk mengimplementasikan metode *K-Nearest Neighbor* dalam memprediksi penjualan produk obat-obatan PT XYZ Palembang menggunakan data penjualan pada periode sebelumnya. Selain itu, peneliti ingin mengevaluasi performa atau tingkat akurasi dari metode K-NN terhadap hasil prediksi penjualan produk obat-obatan. Penelitian ini juga dapat membantu perusahaan dalam mengoptimalkan perencanaan stok obat, sehingga dapat mengurangi risiko kelebihan atau kekurangan ketersediaan obat.

Pemilihan metode *K-Nearest Neighbor* didasarkan pada beberapa keunggulan, seperti ketahanan terhadap data pelatihan yang memiliki gangguan (*noise*) serta tidak memerlukan *preprocessing* data yang rumit (Nuraini Sukmana et al., 2020).

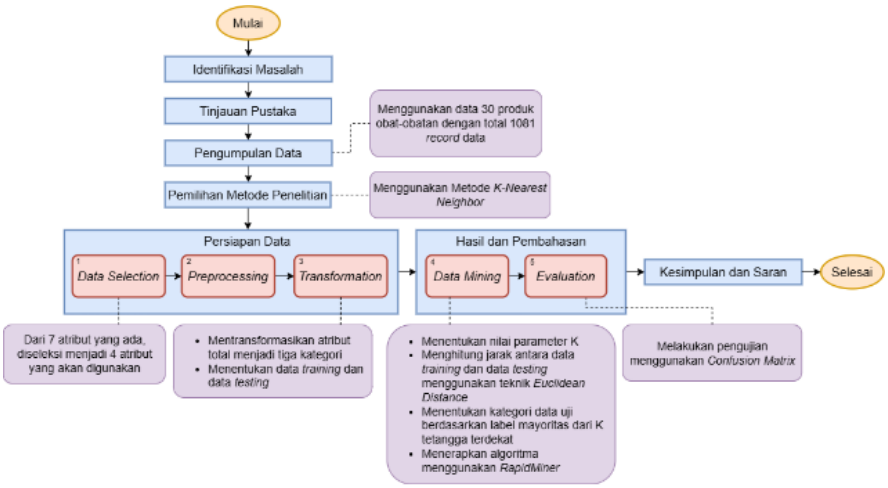
2. METODE PENELITIAN

2.1 Pengumpulan Data

Jenis data yang akan digunakan untuk penelitian ini adalah data primer, yaitu data yang dikumpulkan dan diolah langsung oleh PT XYZ Palembang. Data yang digunakan adalah data mengenai penjualan produk obat-obatan selama tiga tahun terakhir mulai dari tahun 2021 hingga tahun 2023 dengan total 1081 *record* data. Pengumpulan data primer diperlukan untuk bahan penelitian prediksi penjualan berdasarkan data historis penjualan.

Dari data yang telah dikumpulkan, peneliti akan memanfaatkan sejumlah atribut yang relevan sebagai landasan untuk melaksanakan proses prediksi dengan menerapkan algoritma *K-Nearest Neighbor*.

2.2 Tahapan Penelitian



Gambar 1. Kerangka Kerja Penelitian

Gambar di atas merupakan diagram alur tahapan pada proses penelitian ini. Proses *data mining* yang dilaksanakan untuk penelitian ini mengikuti langkah-langkah yang ada di dalam kerangka *Knowledge Discovery in Database* (KDD) untuk memperoleh informasi berdasarkan urutan tahapan yang telah ditetapkan, berikut adalah rincian tahapannya (Bahtiar, 2023).

2.2.1 *Data selection*

Tahapan data *selection* melibatkan proses identifikasi dan memilih data yang relevan dari data operasional yang telah dikumpulkan sebelum ekstraksi informasi dimulai dalam kerangka KDD. Data yang terpilih pada tahap ini akan dimanfaatkan dalam proses *data mining* berikutnya. Fungsi dari data

https://doi.org/10.35145/joisie.v8i2.4725

selection adalah untuk memilih data yang relevan guna mendukung pengolahan dan analisis pada penelitian ini.

2.2.2 *Preprocessing*

Setelah data *selection* selesai, tahap berikutnya adalah *preprocessing* pada data yang telah diperoleh. Proses pembersihan data (*data cleaning*) dilakukan pada tahap ini untuk memastikan bahwa dataset yang dihasilkan bersih dan siap digunakan untuk tahap berikutnya. Proses data *cleaning* melibatkan penghapusan data yang tidak sesuai (relevan), nilai yang tidak terisi (*missing value*), dan data *redundant* karena ini merupakan langkah awal yang penting sebelum melaksanakan *data mining*. *Missing value* terjadi ketika ditemukan atribut dalam kumpulan data yang tidak terisi atau dibiarkan kosong, sementara jika terdapat lebih dari satu *record* data memiliki nilai yang sama dalam dataset maka data dianggap *redundant*.

2.2.3 *Transformation*

Tahap transformasi adalah proses untuk mengubah format data yang telah diseleksi agar siap untuk diterapkan dalam proses *data mining*. Pada tahapan KDD, transformasi adalah langkah inovatif yang bergantung pada tipe data atau pola informasi yang ingin ditemukan dalam basis data. Tujuan dari tahap ini adalah untuk mentransformasikan dan mengkonsolidasi format data yang berbeda yang akan dimanfaatkan dalam proses *data mining*. Transformasi data mencakup aktivitas seperti normalisasi, agregasi (ringkasan), dan generalisasi data.

2.2.4 *Data mining*

Pada tahap ini, fokusnya adalah mengidentifikasi pola atau data informasi yang menarik dalam kumpulan data yang telah dipilih dengan memanfaatkan metode atau teknik tertentu yang sesuai dengan kerangka umum proses KDD. Penelitian ini menerapkan satu dari beberapa metode atau pendekatan dalam proses *data mining* yaitu *K-Nearest Neighbor*. Peneliti melakukan pengolahan data dengan menggunakan perangkat lunak *RapidMiner*, di mana data yang telah siap akan diolah mengikuti desain proses yang dirancang pada *RapidMiner*.

2.2.5 *Evaluation*

Hasil prediksi yang diperoleh dari algoritma *K-Nearest Neighbor* perlu melalui tahap evaluasi untuk membandingkan hasil prediksi yang diberikan oleh model dengan data aktual yang ada. Pengukuran tingkat kesalahan pada penelitian ini menggunakan *Confusion Matrix*.

2.3 *K-Nearest Neighbor* (K-NN)

K-Nearest Neighbor (KNN) adalah sebuah metode atau algoritma yang digunakan untuk mengklasifikasikan objek baru yang didasarkan pada kedekatan antara objek tersebut dengan objek-objek dalam data pelatihan (*training*) (Susilo et al., 2024). Algoritma ini mencari sejumlah tetangga terdekat (K) dalam data pelatihan yang memiliki kemiripan atau kedekatan terhadap objek di dalam data baru (*testing*) (Nuraini Sukmana et al., 2020). K-NN adalah algoritma dalam kategori *supervised learning*, di mana klasifikasi suatu objek baru ditetapkan berdasarkan label kelas yang dominan dari tetangga-tetangga terdekatnya (*nearest neighbor*) (Dewi et al., 2022).

Algoritma *K-Nearest Neighbor* (K-NN) memiliki tujuan untuk melakukan pengelompokan objek baru dengan memanfaatkan atribut dan data pelatihan (*training*) yang tersedia (Nijunnihayah et al., 2024). Berikut ini adalah langkah-langkah perhitungan dalam metode *K-Nearest Neighbor* (Nijunnihayah et al., 2024).

1) Menetapkan nilai K yang akan digunakan sebagai jumlah tetangga paling dekat dalam perhitungan K-NN ini.
2) Mengukur jarak antara data pengujian (*testing*) terhadap semua data pelatihan (*training*) dengan memanfaatkan teknik perhitungan *Euclidean Distance*. Rumus untuk menentukan *Euclidean Distance* ditunjukkan pada persamaan 1 berikut.

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

https://doi.org/10.35145/joisie.v8i2.4725

Keterangan:
d_i = *Euclidean Distance*
x_i = nilai x_i pada data pelatihan (*training*)
y_i = nilai y_i pada data pengujian (*testing*)
n = jumlah banyaknya data

3) Mengurutkan data yang telah dilakukan perhitungan tersebut ke dalam kelompok yang memiliki jarak *Euclidean* paling kecil.
4) Menentukan kelas label atau kategori berdasarkan nilai K tetangga terdekatnya.
5) Dengan memanfaatkan kategori tetangga terdekat (*nearest neighbor*) yang paling dominan atau mayoritas, maka objek baru dapat diprediksikan.

3. HASIL DAN PEMBAHASAN

Pembahasan berikut ini akan diuraikan mengenai tahapan dalam pengolahan *data mining* melalui penerapan algoritma *K-Nearest Neighbor* untuk memperkirakan penjualan obat, serta menyajikan informasi berdasarkan urutan langkah yang telah ditetapkan. Berikut adalah langkah-langkahnya.

3.1 Tahapan *Data Selection*

Tahapan pemilihan atau penyaringan pada data dilakukan untuk memperoleh data yang relevan dengan kebutuhan dalam memprediksi penjualan obat-obatan. Seluruh atribut yang terdapat dalam data penjualan meliputi *field* Tahun, Bulan, *Lini Desc*, *Sales QTY*, *Branch Desc*, Kode Obat, dan Zat Aktif. Untuk melakukan prediksi penjualan dibutuhkan beberapa atribut, maka dari itu seluruh atribut yang ada akan diseleksi menjadi 4 atribut yaitu:

1) Tahun, merupakan atribut yang berisi informasi mengenai tahun dari transaksi penjualan obat-obatan.
2) Bulan, merupakan atribut yang berisi informasi mengenai periode bulan dari transaksi penjualan obat-obatan.
3) Kode Obat, merupakan atribut yang berisi informasi mengenai kode produk obat-obatan yang dijual oleh PT XYZ Palembang.
4) *Sales QTY*, merupkan atribut yang berisi informasi mengenai jumlah transaksi penjualan obat-obatan pada satu periode (bulan).

Berikut ini adalah hasil *data selection* yang dipakai dari data mentah yang ada:

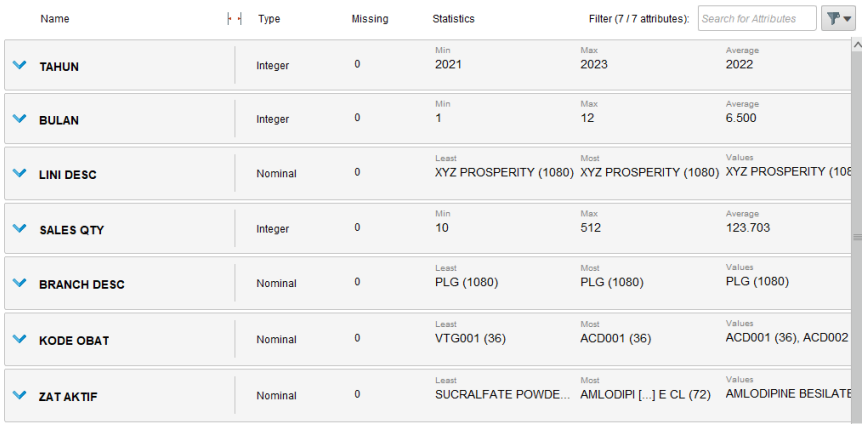
Tabel 1. Hasil *Data Selection*

No	Tahun	Bulan	Kode Obat	Sales QTY
1	2021	1	BCR001	68
2	2021	2	BCR001	42
3	2021	3	BCR001	23
4	2021	4	BCR001	51
5	2021	5	BCR001	84
6	2021	6	BCR001	110
7	2021	7	BCR001	77
…	…	…	…	…
1081	2023	12	ANTM002	272

3.2 Tahapan *Preprocessing*

Tahap berikutnya dalam KDD adalah tahap *preprocessing*, yang perlu dilaksanakan sebelum proses *data mining* dimulai. Tahapan persiapan atau pembersihan data ini dilakukan sebelum data tersebut digunakan dalam algoritma K-NN untuk klasifikasi. Karena K-NN sangat bergantung pada jarak antar data, penting untuk memastikan data memiliki format yang tepat dan tidak mengandung *outlier* atau *noise* yang berlebihan, yang dapat mempengaruhi hasil akhir. Dalam tahapan ini tidak ada penggabungan atau pembersihan data karena data tidak memiliki *missing value* atau bernilai 0 (nol), sehingga memungkinkan untuk dapat melanjutkan ke tahap berikutnya.

https://doi.org/10.35145/joisie.v8i2.4725



Gambar 2. Hasil *Preprocessing* data

3.3 Tahapan *Transformation*

Tahap *transformation* berfungsi untuk mentransformasikan format data yang akan diterapkan dalam proses *data mining*. Tahap *transformation* meliputi normalisasi, agregasi, atau generalisasi data. Dari hasil seleksi sebelumnya, seluruh data penjualan obat-obatan dalam tahapan ini akan menghasilkan pengelompokan atribut, yaitu atribut dari data bulan. Pada tahap transformasi ini, data penjualan obat dikelompokkan menjadi tiga kategori yang membutuhkan total penjualan tiap bulannya untuk mempermudah dalam prediksi penjualan. Berikut adalah seluruh data yang telah dikelompokkan dan diurutkan.

Tabel 2. Data Penjualan Sebelum Ditransformasikan

No	Kode Obat	Jan	Feb	Mar	…	Des	Total
1	ACD001	343	204	194	…	321	2976
2	ACD002	223	188	184	…	215	2561
3	ACD004	115	103	181	…	178	2033
4	ANTM001	476	355	484	…	196	3682
5	BCK001	161	172	120	…	163	1936
6	DBT001	386	406	351	…	247	3647
7	CAL001	367	208	379	…	236	3049
8	ATH001	121	123	134	…	128	1522
9	ATH002	135	183	304	…	200	2363
10	ATH003	162	167	171	…	184	1975
…	…	…	…	…	…	….	…
30	RTZ001	178	164	168	…	135	2014

Berdasarkan data diatas, transformasi akan dilakukan dengan mengubah atribut total menjadi tiga kategori yaitu Sangat Laris, Cukup Laris, dan Kurang Laris. Tahapannya adalah sebagai berikut (Bahtiar, 2023):

a) Menetapkan nilai maksimal dan minimal dari data total yang terdapat pada tabel di atas.
   Data total maksimal = 3861
   Data total minimal = 1522
b) Menghitung selisih antara data total maksimal dan data total minimal.
   Data total maksimal – data total minimal = 3861 – 1522 = 2399
c) Menghitung rentang dengan cara membagi hasil selisih operasi sebelumnya dengan jumlah kategori.
   Rentang = (data total maksimal – data total minimal) ÷ 3
   Rentang = 2339 ÷ 3 = 780
d) Menetapkan Kategori 1 dengan menjumlahkan rentang dan data total minimal.
   Kategori 1 = rentang + total minimal
   Kategori 1 = 780 + 1522 = 2302 (Kurang Laris)
e) Menetapkan Kategori 2 dengan menjumlahkan rentang dan kategori 1.
   Kategori 2 = rentang + Kategori 1
   Kategori 2 = 780 + 2302 = 3082 (Cukup Laris)

https://doi.org/10.35145/joisie.v8i2.4725

f) Menetapkan Kategori 3 dengan menjumlahkan rentang dan kategori 2.
Kategori 3 = rentang + Kategori 2
Kategori 3 = 780 + 3082 = 3862 (Sangat Laris)

Sehingga, diperoleh tabel Kategori seperti berikut:

Tabel 3. Tabel Kategori

Total	Kategori
≤ 2302	Kurang Laris
$> 2302 \leq 3082$	Cukup Laris
$> 3082 \leq 3862$	Sangat Laris

Berikut adalah data penjualan yang telah ditransformasikan berdasarkan kategori.

Tabel 4. Data Penjualan Setelah Ditransformasikan

No	Kode Obat	Jan	Feb	Mar	…	Des	Kat
1	ACD001	343	204	194	…	321	Cukup Laris
2	ACD002	223	188	184	…	215	Cukup Laris
3	ACD004	115	103	181	…	178	Kurang Laris
4	ANTM001	476	355	484	…	196	Sangat Laris
5	BCK001	161	172	120	…	163	Kurang Laris
6	DBT001	386	406	351	…	247	Sangat Laris
7	CAL001	367	208	379	…	236	Cukup Laris
8	ATH001	121	123	134	…	128	Kurang Laris
9	ATH002	135	183	304	…	200	Cukup Laris
10	ATH003	162	167	171	…	184	Kurang Laris
…	…	…	…	…	…	….	…
30	RTZ001	178	164	168	…	135	Kurang Laris

Data yang telah di transformasikan tersebut akan dipecah menjadi dua kelompok data yang akan dibutuhkan dalam pemodelan menggunakan *RapidMiner* yaitu data pelatihan (*training*) dan data pengujian (*testing*) sebagai validasi atau memastikan bahwa model yang dihasilkan baik dan dapat bekerja dengan efektif pada kumpulan data yang lainnya. Adapun pembagian antara data pelatihan dan data pengujian dilakukan dengan menetapkan satu tahun terakhir, yaitu tahun 2023, sebagai data pengujian (*testing*), sedangkan dua tahun sebelumnya yaitu tahun 2021 dan tahun 2022, akan digunakan sebagai data pelatihan (*training*), dengan masing-masing berjumlah 20 data *training* dan 10 data *testing*. Di bawah ini terdapat tabel yang menyajikan data pelatihan (*training*) dan data pengujian (*testing*).

Tabel 5. Data *Training*

No	Kode Obat	Jan	Feb	Mar	…	Des	Kat
1	ACD001	343	204	194	…	321	Cukup Laris
2	ACD002	223	188	184	…	215	Cukup Laris
3	ACD004	115	103	181	…	178	Kurang Laris

https://doi.org/10.35145/joisie.v8i2.4725

4	ANTM001	476	355	484	…	196	Sangat Laris
5	BCK001	161	172	120	…	163	Kurang Laris
6	DBT001	386	406	351	…	247	Sangat Laris
7	CAL001	367	208	379	…	236	Cukup Laris
8	ATH001	121	123	134	…	128	Kurang Laris
9	ATH002	135	183	304	…	200	Cukup Laris
10	ATH003	162	167	171	…	184	Kurang Laris
…	…	…	…	…	…	….	….
20	PCH001	298	233	290	…	205	Cukup Laris

Tabel 6. Data *Testing*

No	Kode Obat	Jan	Feb	Mar	…	Des	Kat
1	ACD003	118	142	344	…	431	?
2	ANTM002	230	203	258	…	272	?
3	DBT002	323	382	322	…	279	?
4	CAL002	352	294	282	…	309	?
5	ATH004	196	178	177	…	179	?
6	URG002	124	160	126	…	166	?
7	BCR003	158	147	149	…	120	?
8	RHM003	142	132	150	…	158	?
9	VTG001	343	340	241	…	262	?
10	RTZ001	178	164	168	…	135	?

3.4 Tahapan *Data Mining*

Tahapan *data mining* merupakan tahapan implementasi model prediksi menggunakan *K-Nearest Neighbor*. Pada tahap *data mining*, peneliti mengidentifikasi pola dan informasi yang menarik pada data yang telah dipilih dengan menerapkan metode K-NN sesuai dengan keseluruhan proses KDD. Dalam lingkup penelitian ini, algoritma K-NN dimanfaatkan untuk memprediksi produk obat berdasarkan pada data terdekat yang akan diklasifikasikan dan menentukan label atau kelas data tersebut berdasarkan mayoritas label data terdekat. Berikut adalah rangkaian tahapan perhitungan data menggunakan metode *K-Nearest Neighbor* dengan memanfaatkan data pelatihan dan data uji. Tahapan-tahapan ini meliputi:

a) Menetapkan nilai K. Karena pemilihan nilai K dapat ditentukan secara bebas, maka dalam penelitian ini penulis memilih nilai K adalah 5.
b) Melakukan perhitungan jarak antara data pelatihan (*training*) dan data uji (*testing*) menggunakan *Euclidean Distance* sebagai metode perhitungan dalam penelitian ini. Perhitungannya adalah sebagai berikut.

$$D_{1.1} = \sqrt{(343-118)^2 + (204-142)^2 + (194-344)^2 + (221-242)^2 + (242-219)^2 + (234-113)^2 + (274-178)^2 + (228-118)^2 + (204-367)^2 + (138-171)^2 + (373-337)^2 + (321-431)^2}$$

$$D_{1.1} = 337{,}52$$

Berikut adalah hasil perhitungan jarak menggunakan teknik *Euclidean Distance* antara data pelatihan (*training*) dan data uji (*testing*).

https://doi.org/10.35145/joisie.v8i2.4725

Tabel 7. Hasil Perhitungan Jarak *Euclidean*

No	D1	D2	D3	D4	…	D10
1	337,52	310,41	384,69	359,73	…	388,78
2	415,07	292,86	405,92	378,36	…	188,74
3	407,57	397,28	551,30	502,84	…	143,17
4	660,65	435,82	407,42	450,25	…	606,82
5	458,73	435,75	555,62	526,68	…	107,78
6	536,26	446,93	279,23	314,94	…	580,11
7	475,50	305,44	374,13	381,10	…	383,27
8	522,02	551,89	666,03	627,85	…	157,77
9	411,48	398,69	473,11	468,75	…	215,74
10	429,92	453,15	544,00	502,39	…	110,07
…	…	…	…	…	…	…
20	401,36	250,03	289,82	285,74	…	329,63

c) Data yang telah dihitung akan diurutkan berdasarkan hasil perhitungan jarak mulai dari jarak yang terdekat sampai jarak yang terjauh (*ascending*).

Tabel 8. Urutan Data Hasil Perhitungan

No Urut	No D1	D1	No D2	D2	…	No D10	D10
1	1	377,52	3	250,03	…	12	66,35
2	19	387,49	8	292,86	…	5	107,78
3	20	401,36	11	305,44	…	10	110,07
4	3	407,57	1	310,41	…	3	143,17
5	9	411,48	19	310,88	…	15	145,53
6	2	415,07	4	312,11	…	17	145,78
7	11	422,82	2	397,28	…	14	146,62
8	12	424,28	7	398,69	…	11	146,68
9	10	429,92	6	406,33	…	8	157,77
10	5	458,73	5	421,58	…	13	185,37
…	…	…	…	…	…	…	…
20	4	660,65	16	551,89	…	4	606,82

d) Langkah selanjutnya yaitu menentukan kategori data hasil uji dengan melihat label yang paling banyak muncul dari K tetangga terdekat. Dengan nilai K=5, maka dipilih 5 tetangga terdekat (*nearest neighbor*) berdasarkan hasil data yang telah diurutkan sesuai data *testing* yang berjumlah 10 data.

e) Data *testing* akan diurutkan berdasarkan mayoritas pada masing-masing data, dengan menetapkan kategori sesuai dengan hasil prediksi yang telah dilakukan perhitungan secara manual.

Berikut adalah hasil untuk proses yang telah dijalankan.

Tabel 9. Hasil Prediksi Data

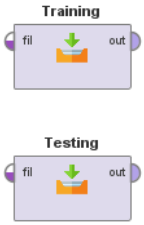
No	Kode Obat	Jan	Feb	Mar	…	Des	Kat
1	ACD003	118	142	344	…	431	Cukup Laris
2	ANTM002	230	203	258	…	272	Kurang Laris
3	DBT002	323	382	322	…	279	Sangat Laris
4	CAL002	352	294	282	…	309	Sangat Laris
5	ATH004	196	178	177	…	179	Kurang Laris
6	URG002	124	160	126	…	166	Kurang Laris
7	BCR003	158	147	149	…	120	Kurang Laris

https://doi.org/10.35145/joisie.v8i2.4725

8	RHM003	142	132	150	…	158	Kurang Laris
9	VTG001	343	340	241	…	262	Sangat Laris
10	RTZ001	178	164	168	…	135	Kurang Laris

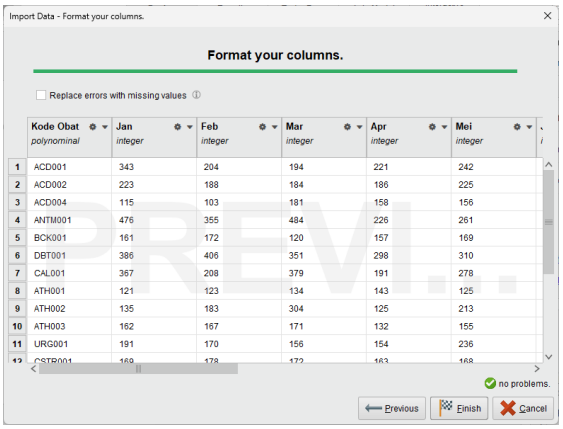
Setelah sebelumnya melakukan proses prediksi secara manual, yaitu dengan melakukan prediksi penjualan menggunakan data penjualan dari tahun 2021 sampai tahun 2023. Selanjutnya peneliti akan mengaplikasikan algoritma *K-Nearest Neighbor* menggunakan perangkat lunak *RapidMiner*.

Tahapan ini dimulai dengan menyiapkan data pelatihan (*training*) dan data uji (*testing*). Peneliti membagi data *training* sebanyak 20 data penjualan dari tahun 2021 sampai tahun 2022 dan data *testing* sebanyak 10 data penjualan dari tahun 2023.



Gambar 3. Membagi Data *Training* dan Data *Testing*

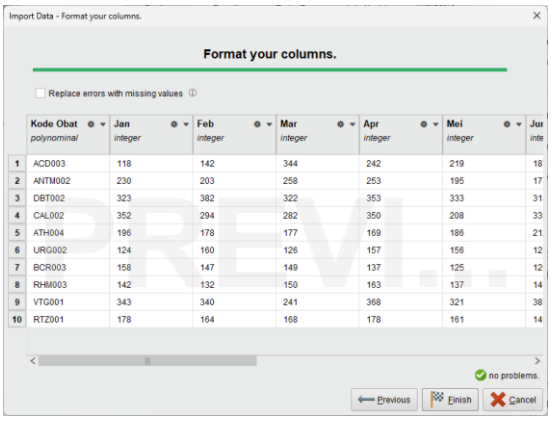
Setelah menyiapkan data *training* dan data *testing*, lakukan *drag and drop* pada operator *read excel* untuk melakukan *importing* tabel *Microsoft Excel* ke dalam proses, sehingga dapat mengekstrak data *training* dan data *testing* dari format excel.



Gambar 4. *Importing* Data *Training*

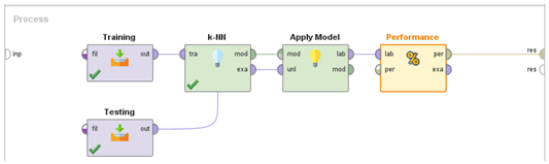
Pada kegiatan *import* data *training*, berisi data penjualan 20 produk obat yang sebelumnya telah diakumulasikan dan dijumlahkan setiap bulan pada tahun 2021 dan 2022. Terdapat 14 variabel diantaranya variabel Kode Obat dengan tipe *polynomial*, variabel bulan Januari hingga Desember dengan tipe *integer*, dan variabel Kategori dengan tipe *polynomial*. Kemudian *set role* atribut Kategori menjadi label yang akan digunakan sebagai target atau variabel yang ingin diprediksi oleh model.

https://doi.org/10.35145/joisie.v8i2.4725



Gambar 5. *Importing* Data *Testing*

Pada kegiatan *import* data *testing*, berisi data penjualan 10 produk obat pada tahun 2023 yang telah dilakukan perhitungan manual untuk menentukan kategori dari masing-masing produk. Sama halnya seperti pada data *training*, terdapat 14 variabel diantaranya variabel Kode Obat dengan tipe *polynomial*, variabel bulan Januari hingga Desember dengan tipe *integer*, dan variabel Kategori dengan tipe *polynomial*. Kemudian *set role* atribut Kategori menjadi label yang akan digunakan sebagai target atau variabel yang ingin diprediksi oleh model.



Gambar 6. Susunan Operator Model K-NN

Kemudian, lakukan *drag and drop* pada operator K-NN dengan memasukkan nilai K=5, lalu sambungkan operator *read excel* (*training*) dengan operator K-NN. Adapun fungsi dari operator K-NN yaitu sebagai representasi proses algoritma K-NN itu sendiri. Kemudian, ulangi *drag and drop* pada operator *apply model* lalu sambungkan operator K-NN dan data *testing* pada *apply model*. Setelah model K-NN diterapkan, langkah terakhir adalah mengukur performa model menggunakan operator *performance* (*classification*). Operator ini digunakan untuk menentukan tingkat akurasi dari kinerja model K-NN yang dilakukan dengan cara membandingkan hasil prediksi model dengan nilai aktual pada dataset uji.

KNNClassification

```
Weighted 5-Nearest Neighbour model for classification.
The model contains 20 examples with 13 dimensions of the following classes:
  Cukup Laris
  Kurang Laris
  Sangat Laris
```

Gambar 7. Klasifikasi K-NN

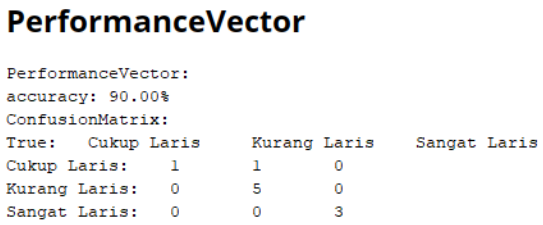
Pada gambar diatas menunjukkan hasil dari model klasifikasi atau pengelompokan dengan menggunakan algoritma *K-Nearest Neighbor* (KNN). Dalam hal ini, model KNN menggunakan 5 tetangga terdekat dan menerapkan pembobotan pada tetangga terdekat tersebut untuk menentukan kelas yang dihasilkan. Model ini menggunakan 20 data *training* dan mengindikasikan bahwa setiap dataset memiliki 13 atribut (fitur) yang berbeda. Dimensi ini mengacu pada jumlah kolom dalam dataset kecuali kolom label. Beberapa kelas yang teridentifikasi dengan model yaitu Cukup Laris, Kurang Laris, dan Sangat Laris.

3.5 Tahapan *Evaluation*

Pola informasi yang terbentuk dalam proses *data mining* perlu disajikan dengan format yang mudah dipahami oleh semua pihak yang terlibat. Tahap ini bertujuan untuk mengevaluasi apakah pola

https://doi.org/10.35145/joisie.v8i2.4725

informasi yang diperoleh sesuai atau berlawanan dengan fakta atau hipotesis yang telah ada. Dalam tahap ini, algoritma *K-Nearest Neighbor* diuji dengan menerapkan *Confusion Matrix*. Pengujian ini bertujuan untuk menentukan tingkat akurasi prediksi yang dilakukan sebelumnya. Hasil dari pengujian tersebut adalah sebagai berikut:



Gambar 8. Hasil Tingkat Akurasi

Tabel 10. Hasil Pengujian *Confusion Matrix*

		True			*Class Precision*
		Cukup Laris	Kurang Laris	Sangat Laris	
Pred.	Cukup Laris	1	1	0	50%
	Kurang Laris	0	5	0	100%
	Sangat Laris	0	0	3	100%
Class Recall		100%	83,33%	100%	

Tabel diatas menunjukkan hasil *Confusion Matrix* dengan tingkat akurasi yang dicapai sebesar 90%. Berikut adalah interpretasi untuk setiap kategori.

a. *Class Precision*
   1) Pada prediksi kategori Cukup Laris, dari total 2 data, hasil menunjukkan bahwa 1 data tergolong Cukup Laris, 1 data tergolong Kurang Laris, dan 0 data tergolong Sangat Laris, dengan tingkat *class precision* sebesar 50%.
   2) Pada prediksi kategori Kurang Laris, dari total 5 data, hasil menunjukkan bahwa 5 data tergolong Kurang Laris serta 0 data tergolong Cukup Laris dan Sangat Laris, dengan tingkat *class precision* sebesar 100%.
   3) Pada prediksi kategori Sangat Laris, dari total 3 data, hasil menunjukkan bahwa 3 data tergolong Sangat Laris, serta 0 data tergolong Cukup Laris dan Kurang Laris, dengan tingkat *class precision* sebesar 100%.

b. *Class Recall*
   1) Pada *true* kategori Cukup Laris, dari satu data yang ada, hasil menunjukkan bahwa satu data tergolong Cukup Laris serta 0 data tergolong Kurang Laris dan Sangat Laris, dengan tingkat *class recall* sebesar 100%.
   2) Pada *true* kategori Kurang Laris, dari total 6 data yang ada, hasil menunjukkan bahwa satu data tergolong Cukup Laris, 5 data tergolong Kurang Laris, dan 0 data tergolong Sangat Laris, dengan tingkat *class recall* sebesar 83.33%.
   3) Pada *true* kategori Sangat Laris, dari total 3 data yang ada, hasil menunjukkan bahwa 3 data tergolong Sangat Laris serta 0 data tergolong Cukup Laris dan Kurang Laris, dengan tingkat *class recall* sebesar 100%.

4. SIMPULAN

Pada penelitian ini, perhitungan dilakukan dengan menggunakan 20 data pelatihan (*training*) dan 10 data pengujian (*testing*), menerapkan teknik *Euclidean Distance* dan menggunakan nilai K = 5 untuk menentukan jarak tetangga terdekat. Pengujian algoritma *K-Nearest Neighbor* (K-NN) dalam prediksi penjualan obat-obatan ini menggunakan *Confusion Matrix* memperoleh tingkat akurasi yang baik yaitu sebesar 90% melalui penerapan *software RapidMiner*.

https://doi.org/10.35145/joisie.v8i2.4725

Berdasarkan *output* prediksi yang didapatkan, mengindikasikan bahwa metode K-NN efektif untuk memprediksi penjualan di masa mendatang. Model K-NN dapat menjadi salah satu solusi untuk memprediksi penjualan obat-obatan berdasarkan data penjualan yang telah ada sebelumnya. Adapun saran bagi penelitian selanjutnya yang akan menerapkan metode ini yaitu membandingkan kinerja K-NN dengan metode lainnya seperti *Linear Regression*, Algoritma C4.5, *Naïve* Bayes, untuk mengetahui metode mana yang paling efektif dalam memprediksi penjualan obat. Selain itu, menambahkan variabel tambahan seperti faktor internal maupun eksternal yang dapat mempengaruhi pola penjualan serta meningkatkan akurasi dari hasil prediksi.

5. DAFTAR PUSTAKA

Bahtiar, R., 2023. Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor. Jurnal Informatika MULTI 1.

Dewi, S.P., Nurwati, N., Rahayu, E., 2022. Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. Building of Informatics, Technology and Science (BITS) 3, 639–648. https://doi.org/10.47065/bits.v3i4.1408

Duran, P.A., Vitianingsih, A.V., Riza, Moch.S., Maukar, A.L., Wati, S.F.A., 2024. Data Mining Untuk Prediksi Penjualan Menggunakan Metode Simple Linear Regression. TEKNIKA 13, 27–34. https://doi.org/10.34148/teknika.v13i1.712

Jessfry, V., Siddik, M., 2024. Penerapan Data Mining Menggunakan Algortima Apriori Dalam Membangun Sistem Persediaan Barang. JOISIE Journal Of Information Systems And Informatics Engineering 8, 187–199.

Nangi, J., Muchtar, M., Teknik Informatika, J., Teknik, F., Halu Oleo, U., 2019. Aplikasi Prediksi Penjualan Barang Menggunakan Metode K-Nearest Neighbor (Knn) (Studi Kasus Tumaka Mart). semanTIK : Teknik Informasi 3, 151–160.

Nijunnihayah, U., Hilabi, S.S., Nurapriani, F., Novalia, E., 2024. Implementasi Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes. MALCOM: Indonesian Journal of Machine Learning and Computer Science 4, 695–701. https://doi.org/10.57152/malcom.v4i2.1326

Nuraini Sukmana, R., Wicaksono, Y., Bandung, S., 2020. Implementasi K-Nearest Neighbor Untuk Menentukan Prediksi Penjualan (Studi Kasus : Pt Maksiplus Utama Indonesia), Jurnal Teknologi Informasi dan Komunikasi.

Pohan, D.A., Dar, M.H., Irmayanti, 22AD. Penerapan Data Mining untuk Prediksi Penjualan Produk Sepatu Terlaris Menggunakan Metode Regresi Linier Sederhana. InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan 6, 257–261. https://doi.org/10.30743/infotekjar.v6i2.4795

Putri, A.A., 2021. Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT. Central Brastagi Utama). RESOLUSI : Rekayasa Teknik Informatika dan Informasi 1, 354–361.

Ramadhona Hassolthine, C., Sahara, R., Asy, F., Haq, S.N., Abdullah, S., 2023. Penerapan Algoritma K-Nearest Neighbour (K-Nn) Sebagai Klasifikasi Tweets Pada Twitter. JOISIE Journal Of Information System And Informatics Engineering 7, 358–363.

Riadi, I., Umar, R., Anggara, R., 2024. Prediksi Kelulusan Tepat Waktu Berdasarkan Riwayat Akademik Menggunakan Metode K-Nearest Neighbor. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) 11, 249–256. https://doi.org/10.25126/jtiik.20241127330

Rismala, Ali, I., Rizki Rinaldi, A., 2023. Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris. Jurnal Mahasiswa Teknik Informatika 7.

Silalahi, A.P., Simanullang, G., Hutapea, M.I., 2023. Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita. METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi 7. https://doi.org/10.46880/jmika.Vol7No1.pp144-149

Susilo, B., Ramdhan, N.A., Bachri, O.S., 2024. Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Produk Digital. MALCOM : Indonesian Journal of Machine Learning and Computer Science 4, 1466–1476. https://doi.org/10.57152/malcom.v4i4.1517

Syafana, V., Hilabi, S.S., Novalia, E., Huda, B., 2024. Prediksi Angka Kelahiran dalam Berbagai Kelompok Umur Ibu Menggunakan Metode K-Nearest Neighbor. MALCOM: Indonesian

Journal of Machine Learning and Computer Science 4, 1096–1103. https://doi.org/10.57152/malcom.v4i3.1392

Yolanda, I., Fahmi, H., 2021. Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor. JIKOMSI (Jurnal Ilmu Komputer dan Sistem Informasi) 3, 9.

https://doi.org/10.35145/joisie.v8i2.4725