

## Klasifikasi Pemilihan Jenis Obat untuk Pasien menggunakan Algoritma K-Nearest Neighbor (K-NN)

Syah Rafi Uddin<sup>1</sup>, Yudha Dwi Putra Negara<sup>2</sup>, Doni Abdul Fatah<sup>3</sup>

Program Studi Sistem Informasi, Fakultas Teknik, Universitas Trunojoyo Madura

E-mail korespondensi: <sup>1</sup>[19044110002@student.trunojoyo.ac.id](mailto:19044110002@student.trunojoyo.ac.id), <sup>2</sup>[yudha.putra@trunojoyo.ac.id](mailto:yudha.putra@trunojoyo.ac.id), <sup>3</sup>[doni.fatah@trunojoyo.ac.id](mailto:doni.fatah@trunojoyo.ac.id),

History Artikel

Diterima : 26 November 2024

Disetujui : 19 Februari 2025

Dipublikasikan : 26 April 2025

---

### ***Abstract***

*Selecting the appropriate type of medication for a patient is a crucial aspect of medical practice that significantly impacts the effectiveness of treatment and the safety of the patient. With the continuous advancement of information technology, decision-making in drug selection can now be supported using machine learning methods. This research aims to develop a classification system for drug selection using the K-Nearest Neighbor (K-NN) algorithm. The data utilized in this study include patient age, gender, blood pressure, and drug type. The K-NN algorithm was chosen due to its ability to classify data based on the proximity of patient attributes to previously labeled data. The optimal parameter  $k$  was determined to maximize prediction accuracy. This system is expected to assist medical professionals in making faster and more accurate decisions in drug selection while reducing the risk of medication errors. The data processing in this study involves several stages, including data understanding, data cleaning, exploratory data analysis (EDA), data preparation, and modeling. The K-Nearest Neighbors (K-NN) model with a parameter  $k = 3$  achieved an accuracy rate of 78%, demonstrating the potential of this approach in the conducted data analysis.*

**Keywords:** *Classification, K-NN, Machine Learning*

## ***Abstrak***

Memilih jenis obat yang sesuai dengan kebutuhan pasien merupakan salah satu aspek penting dalam dunia medis yang mempengaruhi efektivitas pengobatan dan keselamatan bagi pasien itu sendiri. Dengan semakin berkembangnya dunia teknologi informasi, pengambilan keputusan dalam pemilihan obat kini dapat didukung dengan menggunakan metode machine learning. Penelitian ini tentunya juga bertujuan untuk mengembangkan sistem klasifikasi pemilihan jenis obat menggunakan algoritma *K-Nearest Neighbor (K-NN)*. Data yang digunakan dalam penelitian ini terdiri atas, umur pasien, jenis kelamin pasien, tekanan darah pasien, dan juga jenis obat. Algoritma *K-NN* dipilih karena kemampuannya dalam mengklasifikasikan data berdasarkan kedekatan atribut pasien dengan data yang telah terlabel sebelumnya. Parameter  $k$  yang optimal ditentukan untuk memaksimalkan akurasi prediksi. Sistem ini tentunya sangat diharapkan untuk dapat membantu para tenaga medis dalam membuat keputusan yang lebih cepat dan tepat dalam pemilihan obat. Serta mengurangi resiko kesalahan dalam pemberian obat. Proses pengolahan data dalam penelitian ini mencakup tahapan *data understanding, data cleaning, exploratory data analysis (EDA), data preparation, hingga modeling. Model K-Nearest Neighbors (K-NN)* dengan parameter  $k = 3$  berhasil mencapai tingkat akurasi sebesar 78%, menunjukkan potensi pendekatan ini dalam analisis data yang dilakukan.

**Kata Kunci:** *Klasifikasi, K-NN, Machine Learning*

**How to Cite:** Syah Rafi Uddin, Yudha Dwi Putra Negara, Doni Abdul Fatah (2024). Klasifikasi Pemilihan Jenis Obat Untuk Pasien Menggunakan Algoritma K-Nearest Neighbor (K-NN). *KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo*, Vol 9 (1): Halaman 22-31

© 2025 Universitas Muhammadiyah Ponorogo. All rights reserved

---

ISSN 2614-0985 (Print)  
ISSN 2614-0977 (Online)

## PENDAHULUAN

Dalam dunia medis pemilihan jenis obat yang tepat untuk pasien merupakan langkah yang sangat krusial agar bisa memastikan keberhasilan pengobatan. Jika pengambilan keputusan yang kurang tepat, bisa berdampak serius terhadap pasien, seperti efek samping yang dapat merugikan pasien, atau bahkan bisa membahayakan keselamatan pasien itu sendiri. Oleh sebab itu, dibutuhkan nya metode yang efektif dan juga akurat untuk membantu tenaga medis dalam menentukan jenis obat yang sesuai dengan kebutuhan pasien tersebut.

Seiring bertumbuh nya perkembangan teknologi, penerapan kecerdasan buatan dalam bidang kesehatan banyak dilakukan untuk meningkatkan efisiensi dan juga akurasi dalam pengambilan keputusan. Salah satu cara yang dapat dilakukan adalah dengan menggunakan algoritma *Machine Learning*, khususnya algoritma *K-Nearest Neighbor (K-NN)*. Algoritma ini juga di kenal sederhana, namun tentunya sangat efektif dalam menangani berbagai masalah klasifikasi, termasuk juga yang melibatkan data yang kompleks seperti jenis obat yang dibutuhkan oleh pasien..

Algoritma *K-NN* ini bekerja dengan mengklasifikasikan data berdasarkan kedekatan atribut dengan data yang telah terlabel sebelumnya. Didalam konteks pemilihan jenis obat ini, algoritma *K-NN* dapat memanfaatkan berbagai parameter medis, seperti umur pasien, jenis kelamin pasien, tekanan darah pasien, kolesterol pasien, dan jenis obat yang dibutuhkan oleh pasien. Hal itu bertujuan untuk menentukan obat yang paling tepat dan dibutuhkan oleh pasien tersebut. Kemampuan *K-NN* dalam menangani data, dan fleksibilitasnya, menjadikan pilihan yang relevan untuk pengembangan sistem pendukung keputusan dalam dunia medis.

Teknik klasifikasi yang digunakan juga memainkan peran penting dalam menentukan keberhasilan sistem. Metode *K-Nearest Neighbor (KNN)* dikenal

sederhana dan efektif dalam pengenalan pola, serta memanfaatkan fitur warna secara spesifik untuk klasifikasi jenis obat. [1].

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi pemilihan jenis obat untuk pasien menggunakan algoritma *K-NN*. Studi ini tentunya tidak hanya mengevaluasi efektivitas algoritma dalam menentukan obat yang cocok untuk pasien, tetapi juga untuk mengidentifikasi parameter medis yang paling berpengaruh dalam proses klasifikasi. Hal ini tentunya diharapkan dari hasil penelitian ini dapat menjadi kontribusi nyata dalam mendukung pengambilan keputusan medis yang lebih cepat, akurat, dan aman.

## KAJIAN PUSTAKA

Pemilihan jenis obat yang tepat untuk pasien adalah salah satu aspek penting dalam dunia medis untuk memastikan efektivitas pengobatan dan mengurangi risiko efek samping. Dalam era digital, metode berbasis kecerdasan buatan, termasuk algoritma *K-Nearest Neighbor (K-NN)*, telah digunakan untuk membantu proses klasifikasi dalam pengambilan keputusan medis. Tinjauan pustaka ini bertujuan untuk mengkaji teori dasar, penelitian terdahulu, serta relevansi *K-NN* dalam klasifikasi data medis.

*K-Nearest Neighbor (KNN)* metode yang sering digunakan dalam klasifikasi data. Teknik ini digunakan untuk mengelompokkan objek berdasarkan data pembelajaran dengan objek tetangga yang memiliki jarak terdekat atau perbedaan nilai yang kecil [2].

### A. Landasan Teori

Klasifikasi data merupakan cara pengelompokkan benda berdasarkan ciri-ciri yang dimiliki oleh objek. Dalam prosesnya, klasifikasi data dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi. Klasifikasi data yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma komputer. Sedangkan

klasifikasi yang dilakukan dengan bantuan teknologi, yaitu contoh nya dengan menggunakan algoritma *K-Nearest Neighbor (K-NN)* [3].

Dalam konteks penelitian ini, klasifikasi data sering digunakan untuk menganalisis data pasien, seperti gejala, diagnosis, dan rekomendasi obat untuk pasien itu sendiri.

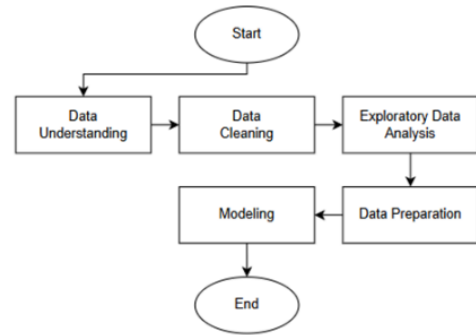
### B. Algoritma K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* adalah algoritma *supervised learning* dimana hasil dari *instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan sampel-sampel dari training data. Algoritma *K-Nearest Neighbor* menggunakan *Neighborhood Classification* sebagai nilai prediksi dari nilai *instance* yang baru [4].

## METODE PENELITIAN

Metode K-Nearest Neighbor adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training sampel. Nilai prediksi dari query ditentukan berdasarkan nilai klasifikasi terdekat. [5]

Dari pengertian tersebut metode K-NN bisa diartikan sebuah metode yang bisa mengklasifikasikan berdasarkan data paling dekat berdasarkan tetangga atau data sebelumnya yang dimiliki sebagai sampel untuk menemukan hasil akhir.



Gambar 1 Tahapan penelitian

Pada gambar 1. Dapat dilihat ada beberapa tahapan yang perlu dilakukan, diantaranya adalah, *data understanding*, *data cleaning*, *exploratory data analysis (EDA)*, *data preparation*, dan *modeling*.

#### 1. Data Understanding

*Data understanding* yaitu memberikan pengertian terkait dengan korelasi dan penyebaran data yang ada sehingga memudahkan untuk melakukan persiapan data [6].

Dalam kata lain untuk kasus rekomendasi obat yang diberikan terhadap pasien itu berdasarkan atribut dari pasien.

#### 2. Data Cleaning

*Data Cleaning* merupakan proses penting yang mencakup dalam identifikasi, perbaikan, dan penghapusan data dari ketidaksesuaian atau anomali dalam suatu *dataset*. Pentingnya *data cleaning* tidak bisa diabaikan karena kualitas visualisasi dan hasil analisis sangat bergantung pada kebersihan dan kualitas data yang digunakan. [7].

#### 3. Exploratory Data Analysis (EDA)

*Exploratory Data Analysis* adalah proses penting untuk memeriksa *dataset* dengan tujuan menemukan pola, mencari hubungan, atau mengidentifikasi kecenderungan yang mungkin tersembunyi di dalam data.

## HASIL DAN PEMBAHASAN

Tujuan utamanya yaitu untuk mendapatkan pemahaman yang lebih luas tentang *dataset* sebelum melakukan analisis yang lebih mendalam atau membangun model prediktif. [8].

### 4. Data Preparation

*Data preparation* dalam *KNN (K-Nearest Neighbors)* adalah proses menyiapkan *dataset* agar dapat digunakan secara optimal oleh algoritma *K-NN*. *Data preparation* sangat penting karena *K-NN* adalah algoritma berbasis jarak, di mana kualitas prediksi sangat bergantung pada struktur dan skala data. Pada tahapan ini data yang sudah didapatkan sebelumnya, diberikan sentimen. Setelah data yang telah didapatkan diberikan sentiment, setelah itu data masuk ke tahap *data preprocessing* [9].

### 5. Modeling

Pada tahapan *modeling*, proses ini dilakukan setelah data melalui proses *splitting* untuk membagi *dataset* menjadi *data training* dan *data testing*. Setelah data terpisah, langkah berikutnya adalah menentukan parameter-parameter yang digunakan dalam metode *machine learning*. Dalam kasus ini, algoritma yang digunakan adalah *K-Nearest Neighbor (K-NN)*, yang memerlukan penetapan nilai *K*, yaitu jumlah tetangga terdekat yang digunakan untuk membuat prediksi. Pada tahapan ini, nilai *K* ditetapkan sebesar 3, yang berarti model mempertimbangkan 3 tetangga terdekat untuk menentukan hasil klasifikasi atau prediksi.

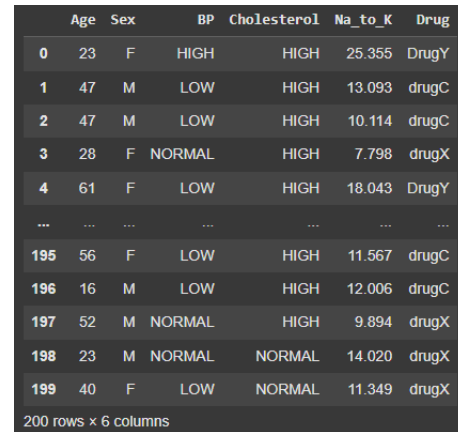
Proses *modeling* meliputi:

#### A. Data training :

*Data training* digunakan untuk membangun model, di mana algoritma akan mempelajari pola-pola dari data.

#### B. Data Testing :

*Data testing* digunakan untuk mengevaluasi kinerja model, seperti akurasi, presisi, *recall*, atau metrik lainnya



|     | Age | Sex | BP     | Cholesterol | Na_to_K | Drug  |
|-----|-----|-----|--------|-------------|---------|-------|
| 0   | 23  | F   | HIGH   | HIGH        | 25.355  | DrugY |
| 1   | 47  | M   | LOW    | HIGH        | 13.093  | drugC |
| 2   | 47  | M   | LOW    | HIGH        | 10.114  | drugC |
| 3   | 28  | F   | NORMAL | HIGH        | 7.798   | drugX |
| 4   | 61  | F   | LOW    | HIGH        | 18.043  | DrugY |
| ... | ... | ... | ...    | ...         | ...     | ...   |
| 195 | 56  | F   | LOW    | HIGH        | 11.567  | drugC |
| 196 | 16  | M   | LOW    | HIGH        | 12.006  | drugC |
| 197 | 52  | M   | NORMAL | HIGH        | 9.894   | drugX |
| 198 | 23  | M   | NORMAL | NORMAL      | 14.020  | drugX |
| 199 | 40  | F   | LOW    | NORMAL      | 11.349  | drugX |

Gambar 2 Data Awal Pasien

Pada gambar 2 dapat dilihat data awal dari pasien yang dilakukan penelitian. Terlihat pada gambar bahwa ada beberapa atribut diantaranya adalah, umur, jenis kelamin, tekanan darah, kolesterol, dan juga jenis obat. Pada gambar 2 juga tertera adanya 200 baris data, dan juga ada 6 kolom data.

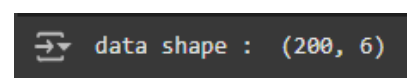
Seperti yang sudah di jelaskan pada bagian metode penelitian, ada 5 tahapan yang dilakukan untuk memproses data tersebut, yaitu :

#### 1. Data Understanding

Pada tahapan *data understanding* ini, penelitian ini melakukan 6 tahapan yang berfungsi untuk mengetahui data yang ada itu secara mendalam, guna bisa memperlancar proses rekomendasi obat yang akan diberikan terhadap pasien berdasarkan atribut dari pasien. Adapun 5 tahapan yang di lakukan yaitu :

##### A. Proses data shape

Pada tahapan proses *data shape* ini dilakukan untuk lebih mengenal terhadap data yang dimiliki.



Gambar 3 Data Shape

Pada gambar 3 dapat dilihat jika *shape* dari data tersebut menampilkan angka 200, dan

6. Angka 200 itu sendiri merepresentasikan sebagai baris dari data, sedangkan angka 6 merepresentasikan kolom dari data tersebut.

B. Menampilkan jenis tipe data

Tahapan menampilkan jenis *type* data ini berfungsi jenis *type* data apa saja yang ada di dalam data tersebut.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Age         200 non-null   int64
1   Sex         200 non-null   object
2   BP          200 non-null   object
3   Cholesterol 200 non-null   object
4   Na_to_K     200 non-null   float64
5   Drug        200 non-null   object
dtypes: float64(1), int64(1), object(4)
```

Gambar 4 Jenis tipe data

Pada gambar 4 dapat dilihat ada 6 atribut yang mana *index* pertama diisi mulai dari 0, yang pertama ada kolom *age*, yang mempunyai *type data integer*. Pada gambar tersebut juga menunjukkan bahwa ada beberapa *type data object* yang nantinya akan di ubah dengan melakukan proses *encoding*, karena untuk menunjang proses pembangunan model, tidak bisa menggunakan *type data object*, maka dari itu akan dilakukan proses *encoding*.

C. Menampilkan *describe* dalam data

|       | Age        | Na_to_K    |
|-------|------------|------------|
| count | 200.000000 | 200.000000 |
| mean  | 44.315000  | 16.084485  |
| std   | 16.544315  | 7.223956   |
| min   | 15.000000  | 6.269000   |
| 25%   | 31.000000  | 10.445500  |
| 50%   | 45.000000  | 13.936500  |
| 75%   | 58.000000  | 19.380000  |
| max   | 74.000000  | 38.247000  |

Gambar 5 Describe data

Pada gambar 5 dapat dilihat beberapa kondisi yang merepresentasikan sebagai jumlah baris pada data. Mean

merepresentasikan nilai rata-rata, std merepresentasikan *standard diffusion*.

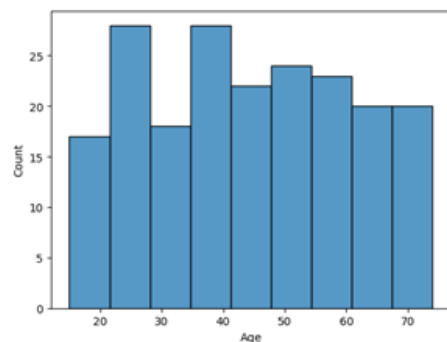
D. Menampilkan nilai yang ada dalam atribut label

| Drug  |    |
|-------|----|
| DrugY | 91 |
| drugX | 54 |
| drugA | 23 |
| drugC | 16 |
| drugB | 16 |

Gambar 6 Atribut label

Pada gambar 6 dapat dilihat tahapan ini difungsikan untuk mengetahui nilai atribut label atau obat yang ada. Pada gambar ditampilkan bahwa atribut obat Y memiliki nilai paling tinggi yaitu sebanyak 91 baris, sedang nilai terkecil ada pada obat C dan B yang memiliki 16 baris. Jadi ketika dijumlahkan keseluruhan data maka hasil akhirnya akan tetap sama sebanyak 200 baris data.

E. Visualisasi atribut umur



Gambar 7 Atribut umur

Pada gambar 7 dapat dilihat terlihat *range* atribut umur dari 15 sampai 75 tahun. Kemunculan umur tertinggi ada dalam umur 25 tahun, dan di 40 tahun.

2. *Cleaning Data*

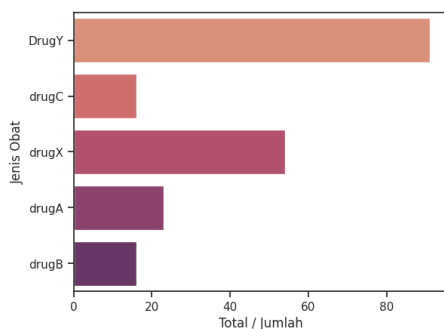
Pada tahapan data *cleaning* ini bertujuan untuk membuang data-data yang tidak konsisten, menghilangkan *noise* dan melengkapi data yang kehilangan informasi, sehingga performansi dari data mining dapat meningkat [10].

Pada tahapan data *cleaning* ini dilakukan 2 tahap, yaitu pembersihan data duplikat, dan *missing value*.

### 3. Exploratory Data Analysis (EDA)

Tahapan selanjutnya setelah *cleaning* data yaitu, proses *exploratory data analysis*, *Exploratory Data Analysis (EDA)* adalah proses analisis terhadap data dengan tujuan memahami karakteristik, pola, dan hubungan di dalam *dataset*. EDA berfokus pada penjabaran dari tiap-tiap atribut yang ada pada dalam data, seperti atribut umur, jenis kelamin akan dilakukan visualisasi untuk masuk ke tahapan *machine learning*. Pada proses *Exploratory Data Analyst* dalam penelitian ini, dibagi dalam 2 proses tahapan, yaitu :

#### A. Visualisasi jenis obat

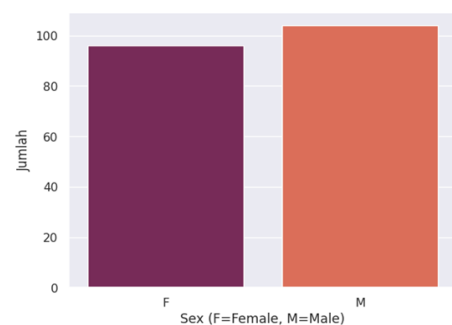


Gambar 8 Jenis obat

Pada gambar 8 dapat dilihat berdasarkan visualisasi yang ditampilkan, dapat diamati bahwa jenis obat dengan frekuensi penggunaan paling tinggi adalah obat Y. Hal ini menunjukkan bahwa obat Y mungkin menjadi pilihan utama dalam pengobatan yang diamati, baik karena efektivitasnya, popularitasnya, atau kondisi medis tertentu yang lebih sering diatasi

menggunakan obat ini. Sebaliknya, frekuensi penggunaan paling rendah ditemukan pada obat C dan obat B. Frekuensi yang rendah ini dapat disebabkan oleh berbagai faktor, seperti kurangnya kebutuhan medis untuk jenis obat tersebut, efek samping yang mungkin lebih tinggi dibandingkan dengan alternatif lain, atau preferensi pengguna yang lebih cenderung memilih obat lain.

#### B. Visualisasi jenis kelamin



Gambar 9 Jenis kelamin

Pada gambar 9 dapat dilihat Sama seperti tahap proses visualisasi sebelumnya, terlihat pada gambar menunjukkan jenis kelamin yang lebih dominan adalah laki-laki dengan jumlah data sebanyak 104, sedangkan untuk perempuan sebanyak 96 data.

### 4. Data Preparation

Pada tahapan data *preparation* ini digunakan untuk melakukan preparasi data seperti melakukan proses *splitting* data ataupun proses *encoding* data. Pada tahapan awal data *preparation*, ada baiknya untuk menampilkan data yang ada terlebih dahulu.

|   | Age | Sex | BP     | Cholesterol | Na_to_K | Drug  |
|---|-----|-----|--------|-------------|---------|-------|
| 0 | 23  | F   | HIGH   | HIGH        | 25.355  | DrugY |
| 1 | 47  | M   | LOW    | HIGH        | 13.093  | drugC |
| 2 | 47  | M   | LOW    | HIGH        | 10.114  | drugC |
| 3 | 28  | F   | NORMAL | HIGH        | 7.798   | drugX |
| 4 | 61  | F   | LOW    | HIGH        | 18.043  | DrugY |

Gambar 10 Data awal pasien

Terlihat pada gambar 10 ada beberapa atribut yang mengandung *type data string* seperti *sex*, *blood pressure*, *cholesterol* dan juga *drug*. Sedangkan untuk proses *machine learning* tidak bisa menggunakan atribut *string*, oleh karena itu, penelitian ini akan mengubah atribut yang mengandung *type data string* menjadi *type data integer* dengan cara melakukan proses *encoding*.

```
from sklearn.preprocessing import LabelEncoder

label_encoder = LabelEncoder()
df['Sex'] = label_encoder.fit_transform(df['Sex'])
df['BP'] = label_encoder.fit_transform(df['BP'])
df['Cholesterol'] = label_encoder.fit_transform(df['Cholesterol'])
df['Drug'] = label_encoder.fit_transform(df['Drug'])
```

Gambar 11 Proses *encoding* data

Terlihat pada gambar 11 atribut yang akan dilakukan proses *encoding* yaitu berupa *atribut sex*, *blood pressure*, *cholesterol*, dan *drug*.

|   | Age | Sex | BP | Cholesterol | Na_to_K | Drug |
|---|-----|-----|----|-------------|---------|------|
| 0 | 23  | 0   | 0  | 0           | 25.355  | 0    |
| 1 | 47  | 1   | 1  | 0           | 13.093  | 3    |
| 2 | 47  | 1   | 1  | 0           | 10.114  | 3    |
| 3 | 28  | 0   | 2  | 0           | 7.798   | 4    |
| 4 | 61  | 0   | 1  | 0           | 18.043  | 0    |

Gambar 12 Data setelah proses *encoding*

Terlihat pada gambar, 12 setelah melakukan proses *encoder*, atribut yang sebelum nya mengandung *type data string*, sudah bisa berubah menjadi *type data integer*. Perubahan *type data* ini sangat penting karena sebagian besar algoritma hanya dapat bekerja dengan data dalam format *type data integer*. Perubahan ini menjadi langkah kunci dalam tahap pra-pemrosesan data untuk menghasilkan model yang lebih akurat dan efisien.

Langkah selanjutnya adalah *proses splitting data* pada semua atribut pada data terkecuali atribut *drug* akan disimpan kedalam *variable X* sedangkan untuk atribut *drug* atau label akan disimpan ke dalam *variable Y*.

```
x : (200, 5)
y : (200,)
```

Gambar 13 *Splitting data*

Terlihat pada gambar 13 dalam variabel *x* menyimpan seluruh atribut terkecuali atribut *drug*, angka 200 merepresentasikan baris data, sedangkan untuk 5 merepresentasikan sebagai atribut yang ada pada dalam data. Didalam variabel *Y* hanya menyimpan atribut *drug* sebanyak 200 baris data.

Tahapan selanjut nya yaitu proses *splitting* data antara data *testing* dan data *training*.

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

Gambar 14 *Splitting data x\_train dan y\_train*

Terlihat pada gambar 14 perintah yang digunakan untuk *splitting* data *testing* dan data *training* dibuat sebanyak 4 *variable*, yaitu *x\_train*, *x\_test*, *y\_train*, dan *y\_test*. Kemudian melakukan *train test splitting* dimana didalam nya ada *variable X* dan juga *Y*. untuk *test size* digunakan untuk membagi data sebanyak 80:20, 80 ada pada *data train*, sedangkan 20 ada pada data *testing*. Kemudian untuk nilai *random state* sebanyak 42.

```
x_train : (160, 5)
x_test : (40, 5)
y_train : (160,)
y_test : (40,)
```

Gambar 15 Setelah proses *splitting data*

Terlihat pada gambar 15 setelah melakukan proses *splitting* data, nilai *variable x* yang sebelumnya ada 200, sekarang menjadi 160, hal itu dikarenakan sudah dibagi untuk nilai *x\_train* hanya diambil dari 80% nilai keseluruhan data.

Sedangkan untuk nilai  $x_{test}$  memiliki 20% dari total keseluruhan data. Berlaku juga pada *variable*  $y_{train}$ , dan  $y_{test}$ .

### 5. Modeling

Langkah pertama pada saat memasuki proses *modeling* yaitu melakukan proses inialisasi model *K-NN* terlebih dahulu dengan cara membuat *variable K-NN* dan untuk nilai parameternya di beri nilai 3. Kemudian setelah inialisasi model *K-NN*, memasukkan *knn.fit* dengan 2 *variable* yaitu  $x_{train}$  dan juga  $y_{train}$ . Kemudian membuat *variable y predict* yang didalam *function predict* ada *variable x test*. Lalu membuat *variable K-NN acc*, didalam *variable K-NN acc* ada *accuracy score* yang bertujuan melihat nilai akurasi pada model yang di bangun.

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 1.00      | 1.00   | 1.00     | 15      |
| 1            | 0.56      | 0.83   | 0.67     | 6       |
| 2            | 0.50      | 0.67   | 0.57     | 3       |
| 3            | 1.00      | 0.20   | 0.33     | 5       |
| 4            | 0.73      | 0.73   | 0.73     | 11      |
| accuracy     |           |        | 0.78     | 40      |
| macro avg    | 0.76      | 0.69   | 0.66     | 40      |
| weighted avg | 0.82      | 0.78   | 0.76     | 40      |

KNN Accuracy : 0.78%

Gambar 16 Hasil *K-NN accuracy*

Terlihat pada gambar 16 bahwa model *K-NN* berhasil mencapai tingkat akurasi sebesar 78%. Nilai ini menunjukkan seberapa baik model *K-NN* dalam mengklasifikasikan data uji dibandingkan dengan label yang sebenarnya. Dengan akurasi sebesar 78%, model *K-NN* dapat dikatakan cukup baik dalam menangani data yang diberikan, meskipun masih ada ruang untuk perbaikan, misalnya dengan menyesuaikan parameter  $k$ , memilih fitur yang lebih relevan, atau melakukan pengolahan data tambahan.

Untuk langkah terakhir yaitu melakukan *testing* pada model yang sudah dibangun dengan membuat *variable testing* yang di dalamnya ada beberapa atribut, terkecuali atribut *drug*. Dari data yang sudah ada, akan menentukan obat apa yang

cocok untuk pasien. Dari *variable testing* tersebut dijadikan *data frame* dan kemudian di tampilkan.

| Age | Sex | BP | Cholesterol | Na_to_K |     |
|-----|-----|----|-------------|---------|-----|
| 0   | 35  | 1  | 2           | 0       | 4.5 |

Gambar 17 *Data testing*

Dapat dilihat dari gambar 17 kurang lebih untuk baris data yang akan dilakukan pengujian seperti itu, data tanpa label atau atribut *drug*.

Kemudian setelah mempunyai data baru tersebut, langkah selanjutnya membuat *variable predict* coba kemudian di dalamnya ada *knn.predict* yang mempunyai *variable testing*.

Hasil Prediksi dari Pasien Baru  
[4]

Gambar 18 Hasil prediksi

Pada gambar 18 terlihat bahwa hasil prediksi model *K-NN* untuk data yang telah diolah sebelumnya adalah bernilai 4. Angka 4 disini secara spesifik untuk merepresentasikan rekomendasi untuk pasien agar menggunakan obat X. Obat X dianggap sebagai pilihan terbaik atau paling sesuai untuk memenuhi kebutuhan atau kondisi yang dihadapi. Prediksi ini mengindikasikan kemampuan model dalam mengenali pola dari data latih dan mengaplikasikannya pada data baru.

## KESIMPULAN DAN SARAN

Penelitian ini berhasil menunjukkan bahwa algoritma *K-Nearest Neighbor (K-NN)* dapat digunakan sebagai metode yang efektif untuk membantu pemilihan jenis obat yang tepat berdasarkan atribut pasien seperti umur, jenis kelamin, tekanan darah, dan kolesterol. Proses pengolahan data yang dilakukan mencakup tahapan *data understanding*, *data cleaning*, *exploratory data analysis (EDA)*, *data preparation*, hingga *modeling*. Dengan parameter  $k$  yang ditetapkan sebesar 3, model *K-NN* mencapai tingkat akurasi sebesar 78%. Nilai ini

mencerminkan performa model yang cukup baik dalam mengklasifikasikan data uji, meskipun masih terdapat ruang untuk peningkatan. Model ini juga dapat menghasilkan rekomendasi obat secara spesifik, seperti hasil prediksi bernilai 4 yang merepresentasikan rekomendasi obat X. Hal ini menunjukkan kemampuan algoritma *K-NN* dalam mengenali pola dari data latih dan mengaplikasikannya pada data baru untuk mendukung pengambilan keputusan medis secara cepat dan tepat.

## DAFTAR PUSTAKA

- [1] N. Neighbor, K. N. N. Berbasis, and F. Warna, "PENGKLASIFIKASIAN DAUN SIRIH DENGAN METODE K-," vol. 3, no. 1, pp. 97–106, 2024.
- [2] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 117–124, 2024, doi: 10.47233/jteksis.v6i1.1078.
- [3] A. P. Wibawa, M. Guntur, A. Purnama, M. Fathony Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, pp. 134–138, 2018.
- [4] C. Kerja and A. Neighbor, "Cara Kerja Algoritma k-Nearest Neighbor ( k-NN ) Apa itu Algoritma k-Nearest Neighbor?," no. February, pp. 0–5, 2019.
- [5] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.
- [6] H. T. A. Simanjuntak, "Deteksi Fraud Pada Klaim Layanan Rumah Sakit Menggunakan Model Neural Network," *J. Appl. Technol. Informatics Indones.*, vol. 1, no. 1, pp. 1–9, 2021, doi: 10.54074/jati.v1i1.30.
- [7] A. Putri, "Pentingnya Data Cleaning Sebelum Visualisasi: Teknik Dan Tips," *Teknologipintar.org*, vol. 4, no. 5, pp. 2024–2025, 2024.
- [8] N. T. M. Sagala and F. Y. Aryatama, "Exploratory Data Analysis (EDA): A Study of Olympic Medallist," *Sistemasi*, vol. 11, no. 3, p. 578, 2022, doi: 10.32520/stmsi.v11i3.1857.
- [9] A. D. Adhi Putra, "Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.
- [10] Y. Andini, J. Tata Hardinata, and Y. P. Purba, "Penerapan Data Mining pada Tata Letak Buku Di Perpustakaan Sintong Bingei Pematangsiantar dengan Metode Apriori," vol. 7, pp. 13–18, 2022, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>