Faktor Exacta

Vol. 18, No. 2, June 2025, pp. 120~128
eISSN: 2502-339X, pISSN: 1979-276X, DOI: https://doi.org/10.30998/faktorexacta.v18i2.28604 

# Penerapan Algoritma K-Nearest Neighbors (KNN) dalam Menentukan Jenis KB Menggunakan Google Colab

**Abdul Hadi[1], Decky Ryansyah[2], Goenawan Brotosaputro[3]**
[1,2,3] Universitas Budi Luhur

**Article Info**

***Article history:***

Received Mar 16, 2025
Revised Jul 03, 2025
Accepted Sept 04, 2025

***Keywords:***

*K-Nearest Neighbors* Algorithm
Contraception
Google Colab
Machine Learning
Recommendation System

**ABSTRACT**

This study aims to apply the *K-Nearest Neighbors* (KNN) algorithm to determine the appropriate contraceptive method (KB) based on demographic data and user characteristics. The main problem faced is the lack of an effective decision support system to assist potential KB users in selecting the most suitable contraceptive method according to individual conditions. Additionally, the selection of contraceptive methods is often done manually by healthcare professionals without the support of predictive technology that could enhance recommendation accuracy. This research was conducted using Google Colab as a data processing platform, utilizing Python libraries such as Pandas, NumPy, and Scikit-learn. The dataset used includes information about KB users, including age, number of children, health history, and personal preferences. The data was pre-processed to handle missing values and normalized to suit the analysis. The KNN model was tested with variations of the k value to find the optimal parameter that yields the highest accuracy. The results showed that the KNN algorithm was able to recommend contraceptive methods with an accuracy of 76% at k = 5. The main finding of this study is that the KNN model can be used as a decision support tool to determine the most appropriate contraceptive method for individuals. This research is expected to support efforts to improve reproductive health services through the utilization of machine learning technology.

***Corresponding Author:***

Universitas Budi Luhur,
Jl. Ciledug Raya No.99, Petukangan Utara, Kec. Pesanggrahan, Kota Jakarta Selatan,
Email: 2411600212@student.budiluhur.ac.id

## 1. PENDAHULUAN

Pemilihan metode kontrasepsi (KB) yang tepat merupakan langkah penting dalam mendukung kesehatan reproduksi dan perencanaan keluarga. Namun, proses pemilihan KB sering kali menghadapi tantangan karena kompleksitas karakteristik individu, seperti usia, jumlah anak, riwayat kesehatan, dan preferensi pribadi. Saat ini, keputusan terkait pemilihan KB umumnya diambil secara manual oleh tenaga medis berdasarkan pengalaman dan pengetahuan umum. Meskipun pendekatan ini dapat memberikan solusi bagi sebagian individu, metode tersebut cenderung kurang akurat dan tidak selalu mempertimbangkan variasi unik setiap individu. Akibatnya, banyak pengguna KB yang mengalami ketidakcocokan dengan metode yang dipilih, yang dapat menyebabkan ketidaknyamanan, efek samping, atau bahkan kegagalan dalam mencegah kehamilan [1]. Oleh karena itu, dibutuhkan sistem yang lebih efektif untuk membantu pengguna dan tenaga medis dalam menentukan jenis KB yang sesuai.

Teknologi machine learning telah menunjukkan potensi besar dalam membantu pengambilan keputusan di berbagai bidang, termasuk kesehatan. Beberapa penelitian sebelumnya telah mencoba memanfaatkan teknologi ini untuk mendukung layanan kesehatan, seperti prediksi penyakit dan analisis data medis. Misalnya, algoritma Decision Tree digunakan untuk klasifikasi penyakit berdasarkan data pasien [2], sementara Support Vector Machine (SVM) diaplikasikan untuk merekomendasikan metode KB [3]. Namun, hasil dari penelitian tersebut masih memiliki keterbatasan, seperti sensitivitas terhadap noise dalam dataset dan kesulitan dalam mengatasi data dengan korelasi tinggi antar fitur. Dalam konteks ini, algoritma *K-Nearest Neighbors* (KNN) menawarkan peluang baru untuk meningkatkan akurasi rekomendasi KB. Algoritma KNN dikenal karena kemampuannya untuk bekerja dengan baik pada dataset kecil hingga menengah serta kemudahan interpretasi

hasilnya [4]. Dengan menggunakan data demografi dan karakteristik pengguna KB, algoritma ini dapat memberikan rekomendasi yang lebih personal dan relevan.

Secara teori, algoritma *K-Nearest Neighbors* (KNN) bekerja dengan cara membandingkan data baru dengan data yang sudah ada, lalu menentukan hasil berdasarkan data yang paling mirip. KNN tidak membutuhkan proses pelatihan yang rumit, melainkan langsung menggunakan data yang tersedia untuk membuat prediksi. Hal ini membuat KNN cocok digunakan pada data yang jumlahnya tidak terlalu besar dan memiliki banyak variasi. Dalam penelitian ini, KNN dipakai karena mampu menyesuaikan rekomendasi metode KB dengan kondisi setiap individu, misalnya umur, jumlah anak, atau riwayat kesehatan. Kurangnya sistem pendukung keputusan yang efektif dalam menentukan jenis KB yang sesuai menjadi permasalahan utama. Selain itu, variasi karakteristik individu seperti usia, jumlah anak, riwayat kesehatan, dan preferensi pribadi membuat pemilihan KB menjadi kompleks. Dengan demikian, dibutuhkan pendekatan yang lebih sistematis dan berbasis data untuk menghasilkan rekomendasi yang lebih tepat guna.

Penelitian ini mengusulkan penerapan algoritma *K-Nearest Neighbors (KNN)* sebagai solusi untuk menentukan jenis kontrasepsi (KB) yang sesuai berdasarkan karakteristik individu. Algoritma KNN dipilih karena kemampuannya untuk bekerja dengan baik pada dataset kecil hingga menengah, serta kemudahan dalam interpretasi hasilnya. Proses implementasi dilakukan menggunakan platform Google Colab, yang memungkinkan analisis data secara efisien dan dapat diakses oleh berbagai kalangan. Dataset yang digunakan mencakup informasi demografi dan kesehatan pengguna KB, seperti usia, jumlah anak, riwayat kesehatan, dan preferensi pribadi. Data tersebut dipreproses untuk meningkatkan kualitas sebelum dilakukan pelatihan model. Dengan pendekatan ini, algoritma KNN dapat memberikan rekomendasi KB yang lebih personal dan akurat dibandingkan metode manual tradisional.

Nilai baru dari penelitian ini terletak pada inovasi penggunaan algoritma KNN untuk mendukung pengambilan keputusan dalam pemilihan metode kontrasepsi. Penelitian ini tidak hanya memberikan solusi teknis, tetapi juga berkontribusi pada pengembangan sistem rekomendasi KB berbasis teknologi machine learning yang dapat diintegrasikan ke dalam layanan kesehatan digital. Selain itu, penelitian ini menawarkan wawasan baru tentang potensi pemanfaatan Google Colab sebagai alat bantu dalam pengolahan data untuk penelitian di bidang kesehatan masyarakat. Dengan pendekatan ini, diharapkan dapat meningkatkan efisiensi dan akurasi dalam pemilihan KB, sehingga mendukung peningkatan kualitas layanan kesehatan reproduksi secara keseluruhan. Inovasi ini juga membuka peluang bagi penelitian lanjutan dalam mengembangkan algoritma yang lebih canggih untuk aplikasi serupa di masa depan.

## 2. METODE

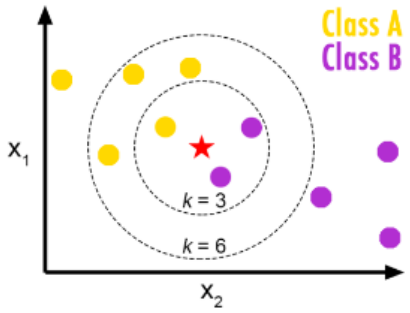
### *2.1 K-Nearest Neighbors (KNN)*

*Nearest Neighbor (KNN)* adalah suatu teknik yang digunakan untuk mengelompokkan objek berdasarkan kesamaan nya dengan objek terdekat [5]. Salah satu algoritma dalam machine learning yang digunakan untuk tugas klasifikasi atau regresi. Algoritma ini termasuk dalam kelompok lazy learning, yang berarti bahwa model tidak memerlukan proses pelatihan eksplisit seperti pada algoritma lainnya, melainkan langsung menggunakan data latih untuk membuat prediksi saat diberikan data baru. KNN bekerja berdasarkan prinsip kemiripan (*similarity*) antara data baru dengan data yang sudah ada dalam dataset. Algoritma ini mencari **k** jumlah tetangga terdekat dari data baru dalam ruang fitur, di mana tetangga terdekat dipilih berdasarkan jarak seperti Euclidean, Manhattan, atau Minkowski. Setelah menemukan **k** tetangga terdekat, algoritma akan memutuskan label atau nilai output untuk data baru berdasarkan mayoritas label dari tetangga tersebut (untuk klasifikasi) atau rata-rata nilai tetangga (untuk regresi). Kelebihan KNN terletak pada kesederhanaan nya, fleksibilitas untuk masalah klasifikasi dan regresi, serta efektivitas nya pada dataset kecil hingga menengah. Namun, algoritma ini memiliki kekurangan seperti sensitivitas terhadap noise dan outlier, kebutuhan normalisasi fitur, serta waktu komputasi yang tinggi pada dataset besar. KNN telah banyak digunakan dalam berbagai bidang, seperti kesehatan untuk klasifikasi penyakit, sistem rekomendasi untuk memberikan saran berdasarkan preferensi pengguna, dan pengenalan pola seperti deteksi objek atau analisis citra. Pada penelitian ini, KNN digunakan untuk merekomendasikan jenis kontrasepsi berdasarkan karakteristik individu, sehingga dapat mendukung pengambilan keputusan yang lebih personal dan akurat dalam layanan kesehatan reproduksi

Metode KNN dibagi menjadi dua fase, yaitu pembelajaran (*training*) dan klasifikasi atau pengujian (*testing*) [6]. Pada babak evaluasi, algoritme ini cuma lakukan penyimpanan vector-vektor feature dan kategorisasi dari data evaluasi. Pada babak kategorisasi, sejumlah fitur yang masih sama dihitung untuk data yang hendak pada uji coba (yang kategorisasi nya tidak dikenali). Jarak dari vector yang baru ini pada semua vector data evaluasi dihitung, dan beberapa k buah neighbor yang terdekat diambil. Penghitungan jarak ketetanggan memakai algoritme euclidien sama seperti yang diperlihatkan pada kesamaan 1

$$\text{euc} = \sqrt{((a_1 - b_1)^2 + \cdots + (a\_n - b\_n)^2)} \qquad (1)$$

Dimana a = a1, a2, .. , an, dan b = b1, b2, .., bn mewakili n nilai atribut dari dua record. Untuk atribut dengan nilai kategori [5]. Sebuah titik akan diprediksi jenisnya berdasarkan pada klasifikasi terbanyak dari neighbor di sekitar nya,ilustrasi nya dapat dilihat pada Gambar 1




Gambar 1. Ilustrasi penggunaan nilai k pada metode KNN

Nilai k yang terbaik untuk KNN bergantung pada data. Pada umumnya, nilai k yang lebih tinggi akan kurangi dampak noise pada kategorisasi, tapi membuat batas di antara tiap kategorisasi jadi lebih kabur. Nilai k yang baik bisa diputuskan optimisasi patokan, contohnya dengan memakai cross-validation. Kasus khusus di mana kategorisasi diprediksi berdasar data evaluasi yang terdekat (dalam kata lain, k = 1) disebutkan algoritme nearest neighbor

Metode KNN dipilih karena sesuai dengan data yang digunakan dalam penelitian. Data pengguna KB memiliki sifat yang beragam, ada yang berupa angka (seperti umur, jumlah anak) dan ada yang berupa kategori (misalnya riwayat kesehatan atau preferensi). KNN dapat menangani kedua jenis data tersebut dengan baik karena hanya menghitung jarak kedekatan antar data. Selain itu, jumlah data penelitian ini relatif sedikit, sehingga KNN lebih efektif dibanding metode lain yang membutuhkan data besar. Dengan cara ini, KNN diharapkan bisa memberikan hasil rekomendasi metode KB yang akurat dan mudah dipahami.

### 2.2 KB (Keluarga Berencana)

KB (Keluarga Berencana) adalah program atau upaya yang bertujuan untuk membantu individu atau pasangan merencanakan jumlah anak dan mengatur jarak kelahiran antara satu anak dengan anak berikutnya [7]. KB juga dikenal sebagai kontrasepsi, yaitu metode atau alat yang digunakan untuk mencegah kehamilan secara sementara atau permanen. Program KB tidak hanya berfokus pada pencegahan kehamilan tetapi juga bertujuan meningkatkan kesehatan reproduksi, kesejahteraan keluarga, dan kualitas hidup. Ada berbagai jenis metode kontrasepsi yang dapat dipilih sesuai dengan kebutuhan, kondisi kesehatan, dan preferensi individu atau pasangan. Metode ini umumnya dibagi menjadi beberapa kategori, seperti metode alami (misalnya kalender, pengukuran suhu basal tubuh), kontrasepsi barrier (seperti kondom, diafragma), kontrasepsi hormonal (pil KB, suntikan, implan), alat kontrasepsi dalam rahim (AKDR/IUD)**,** metode permanen (tubektomi, vasektomi), dan kontrasepsi darurat (pil darurat, AKDR). Pemilihan metode KB harus didasarkan pada faktor-faktor seperti usia, jumlah anak, riwayat kesehatan, gaya hidup, serta preferensi pribadi pasangan. Namun, proses pemilihan ini sering kali menghadapi tantangan karena kompleksitas karakteristik individu yang bervariasi. Oleh karena itu, dibutuhkan pendekatan yang lebih sistematis dan personal. Dalam penelitian ini, teknologi *K-Nearest Neighbors* (KNN) diusulkan sebagai solusi untuk membantu menentukan jenis KB yang paling sesuai berdasarkan data demografi dan karakteristik pengguna. Hal ini bertujuan untuk memberikan rekomendasi yang lebih akurat dan mendukung peningkatan layanan kesehatan reproduksi secara efektif.

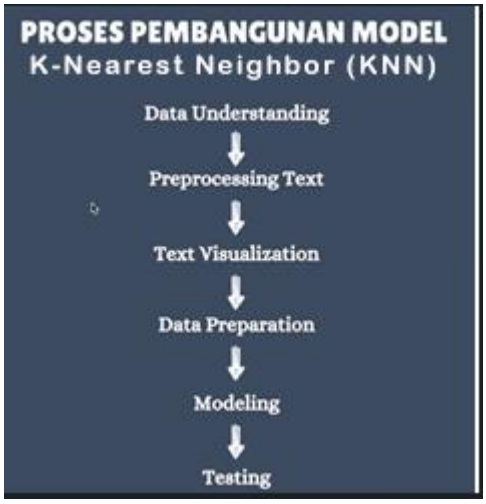
### 2.3 Google Colab

Google Colab adalah platform berbasis cloud yang disediakan oleh Google untuk menjalankan kode pemrograman Python secara online tanpa memerlukan instalasi perangkat lunak tambahan di komputer lokal. Platform ini merupakan bagian dari layanan Google Drive dan dapat diakses melalui browser web, sehingga sangat praktis untuk pengembangan aplikasi machine learning, analisis data, atau eksperimen lainnya yang membutuhkan daya komputasi tinggi [8]. Google Colab menyediakan akses gratis ke lingkungan pengembangan berbasis Jupyter Notebook dengan dukungan GPU (*Graphics Processing Unit*) dan TPU (*Tensor Processing Unit*), yang memungkinkan pengguna untuk menjalankan model machine learning atau deep learning secara efisien. Salah satu keunggulan utamanya adalah kemudahan dalam berbagi dan kolaborasi, karena pengguna dapat menyimpan notebook di Google Drive, mengundang orang lain untuk bekerja sama, atau mempublikasikan hasilnya secara online. Dalam penelitian ini, Google Colab digunakan sebagai platform untuk menerapkan algoritma *K-Nearest Neighbors* (KNN) dalam menentukan jenis kontrasepsi (KB) yang

sesuai berdasarkan karakteristik individu. Dengan menggunakan Google Colab, proses preprocessing data, pelatihan model, dan evaluasi performa dapat dilakukan secara efisien tanpa memerlukan perangkat keras khusus, sekaligus mendukung reproduktifitas penelitian karena kode dan hasil analisis dapat dengan mudah dibagikan kepada pihak lain untuk diverifikasi atau dikembangkan lebih lanjut.

### 2.4 Pengujian Model

Teknik Analisis data menggunakan Data Kuantitatif berupa kaidah-kaidah matematika terhadap anda atau numerik. Analisa dilakuakan melalui algoritma *Nearest Neighbor (KNN)*.Rule yang diperolehdari kedua algoritma tersebut kemudian diuji dengan confusion matrix. Confusion Matrix, ialah langkah tabel untuk mendeskripsikan performa mode perkiraan pada evaluasi supervised learning. Tiap data dari tiap-tiap kelas dalam tabel confusion matrix memperlihatkan jumlah perkiraan yang dibikin buat untuk mengelompokkan kelas yang salah atau benar [9].




Gambar 2. Pengujian Model KNN

## 3. HASIL DAN PEMBAHASAN

Dalam kasus ini,dalam menentukan Jenis KB yang cocok untuk pasien di PMB X sebanyak 254 data pasien, yang terdiri dari 7 atribut. Dimana 6 atribut predictor dan 1 atribut hasil. Penelitian ini bertujuan untuk menentukan jenis KB yang cocok untuk pasien.

| Usia | Jumlah Anak | Kesehatan Reproduksi | Preferensi Durasi | Riwayat Penyakit | Aktivitas Seksual | Jenis KB |
|---|---|---|---|---|---|---|
| 24 | 2 | Sehat | Sementara | Tidak | Jarang | Pil KB |
| 28 | 1 | Sehat | Permanen | Ya | Teratur | IUD |
| 34 | 4 | Gangguan Hormonal | Sementara | Tidak | Teratur | Suntik KB |
| 41 | 4 | Sehat | Permanen | Tidak | Jarang | Tubektomi |
| 30 | 0 | Sehat | Sementara | Tidak | Jarang | Implan |
| 30 | 2 | Gangguan Hormonal | Permanen | Ya | Teratur | Vasektomi |
| 29 | 1 | Sehat | Sementara | Tidak | Teratur | Pil KB |
| 37 | 2 | Sehat | Permanen | Tidak | Jarang | IUD |
| 26 | 0 | Gangguan Hormonal | Sementara | Ya | Jarang | Suntik KB |
| 40 | 5 | Sehat | Permanen | Tidak | Teratur | Tubektomi |
| 27 | 0 | Sehat | Sementara | Tidak | Jarang | Pil KB |
| 34 | 3 | Gangguan Hormonal | Permanen | Ya | Teratur | Vasektomi |
| 31 | 2 | Sehat | Sementara | Tidak | Teratur | Suntik KB |

Gambar 3. Data Mentah

### 3.1 Penyusunan Model Data Mining

Model data mining dalam penelitian ini dibuat dengan menggunakan aplikasi dari google yaitu google colab. Algoritma *Nearest Neighbor (KNN)* dijalankan dengan menggunakan perangkat lunak tersebut. Model yang dihasilkan dijelaskan sebagai berikut.

pembuatan model algoritma *Nearest Neighbor (KNN)* diawali dengan pembacaan file data (Read Excel). Data training disimpan dalam satu file dataset_kb_cleaned .csv.Data ini di simpan dalam google drive,milik peneliti.setelah itu data tersebut di panggil di google Colab dengan cara membuat file baru di Google Colab dengan menambahkan Link yang ada di google Drive

```
# df = pd.read_csv('drug200.csv')
df = pd.read_csv('/content/dataset_kb_cleaned.csv')
df
```

| | Usia | Jumlah_Anak | Kesehatan_Reproduksi | Preferensi_Durasi | Riwayat_Penyakit | Aktivitas_Seksual | Jenis_KB |
|---|---|---|---|---|---|---|---|
| 0 | 24 | 2 | Sehat | Sementara | Tidak | Jarang | Pil KB |
| 1 | 28 | 1 | Sehat | Permanen | Ya | Teratur | IUD |
| 2 | 34 | 4 | Gangguan Hormonal | Sementara | Tidak | Teratur | Suntik KB |
| 3 | 41 | 4 | Sehat | Permanen | Tidak | Jarang | Tubektomi |
| 4 | 30 | 0 | Sehat | Sementara | Tidak | Jarang | Implan |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 249 | 29 | 3 | Gangguan Hormonal | Permanen | Ya | Teratur | Vasektomi |
| 250 | 38 | 4 | Sehat | Permanen | Tidak | Jarang | IUD |
| 251 | 30 | 1 | Gangguan Hormonal | Sementara | Ya | Jarang | Suntik KB |
| 252 | 43 | 5 | Sehat | Permanen | Tidak | Teratur | Tubektomi |
| 253 | 38 | 3 | Gangguan Hormonal | Permanen | Ya | Teratur | Tubektomi |

254 rows × 7 columns

Gambar 4. Import Data dari Google Drive ke Colab

Pada tahap ini, data akan di tampilkan sesuai dengan yang di import dari google drive,sesuai dengan data diatas maka ada data sebanyak 254 baris data dan 7 Kolom.

### 3.2 Data Undestanding

Dalam proses ini di gunakan untuk melihat data-data yang ada pada table

```
[ ] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 254 entries, 0 to 253
Data columns (total 7 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Usia                 254 non-null    int64
 1   Jumlah_Anak          254 non-null    int64
 2   Kesehatan_Reproduksi 254 non-null    object
 3   Preferensi_Durasi    254 non-null    object
 4   Riwayat_Penyakit     254 non-null    object
 5   Aktivitas_Seksual    254 non-null    object
 6   Jenis_KB             254 non-null    object
dtypes: int64(2), object(5)
memory usage: 14.0+ KB
```

Gambar 5. Data Understanding

Dari gambar di atas kita bisa melihat bahwa ada 6 atribut dalam menentukan penggunaan KB. Dan semua memiliki data yaitu integer.

### 3.3 Cleaning Data

*Cleanig Data* adalah proses pembersihan data dari data yang sama atau *duplicate* data agar data yang terbentuk menjadi data yang unik dari data lainnya. Pada tahap ini ,kita akan menghapus data yang *duplicat atua double.*

```
[ ] df.isnull().sum()
```

| | 0 |
|---|---|
| Usia | 0 |
| Jumlah_Anak | 0 |
| Kesehatan_Reproduksi | 0 |
| Preferensi_Durasi | 0 |
| Riwayat_Penyakit | 0 |
| Aktivitas_Seksual | 0 |
| Jenis_KB | 0 |

dtype: int64

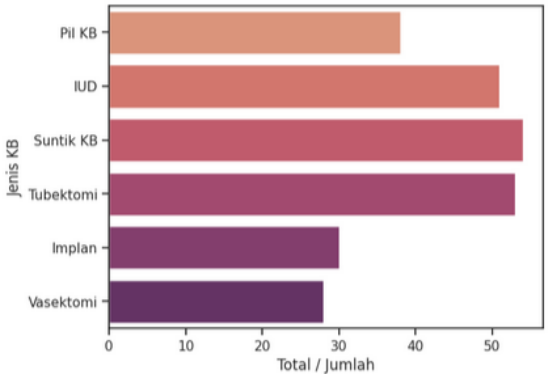
```
[ ] df.duplicated().sum()
0
```

Gambar 6. Cek Duplicate Data

Pada gambar diatas kitab bisa melihat bahwa data tersebut sudah tidak ada lagi data yang duplikat .kita bisa melihat di *df.dublicated().sum()* hasilnya dalah 0 yang mengindentifikasikan bahwa tidak ada data yang sama

### 3.4 *Exploratory Data Analysis (EDA)*

Pada tahap ini,melakukan visualisasi terhadap hasil dari data mining yang di gunakan dalam kasus ini dengan menggunkan diagram batang Batang




Gambar 7. EDA

Dari gambar diatas, menggambarkan bahwa jumlah penggunaan Suntik KB, Tubektomi dan IUD menjadi alternatif penggunaaan kb yang paling diminati. Sedangkan Vasektomi yang paling rendah.

### *3.5 Preparation Data*

*Preparation Data* atau Penyiapan Data adalah tingkatan penting pada proses data mining yang mempunyai tujuan untuk menyiapkan data mentah supaya siap dipakai dalam pemodelan dan analitis. Tingkatan ini meliputi beragam kegiatan seperti pembersihan data (data *cleaning*), integratif data (data *integration*), alih bentuk data (data *transformation*), reduksi data (data *reduction*), dan sample data [10]. Kegiatan pembersihan data mengikutsertakan penghilangan data yang tidak stabil, pembaruan kekeliruan, dan pengatasan *missing values*. Dalam pada itu, integratif data dilaksanakan untuk menyatukan data dari sejumlah sumber jadi satu dataset terintegrasi. Alih bentuk data mencakup normalisasi atau pemetaan feature numerik, perubahan pola data, dan operasi matematika atau statistik pada data.

```
df.head()
```

| | Usia | Jumlah_Anak | Kesehatan_Reproduksi | Preferensi_Durasi | Riwayat_Penyakit | Aktivitas_Seksual | Jenis_KB |
|---|---|---|---|---|---|---|---|
| 0 | 24 | 2 | Sehat | Sementara | Tidak | Jarang | Pil KB |
| 1 | 28 | 1 | Sehat | Permanen | Ya | Teratur | IUD |
| 2 | 34 | 4 | Gangguan Hormonal | Sementara | Tidak | Teratur | Suntik KB |
| 3 | 41 | 4 | Sehat | Permanen | Tidak | Jarang | Tubektomi |
| 4 | 30 | 0 | Sehat | Sementara | Tidak | Jarang | Implan |

| | Usia | Jumlah_Anak | Kesehatan_Reproduksi | Preferensi_Durasi | Riwayat_Penyakit | Aktivitas_Seksual | Jenis_KB |
|---|---|---|---|---|---|---|---|
| 0 | 24 | 2 | 1 | 1 | 0 | 0 | 2 |
| 1 | 28 | 1 | 1 | 0 | 1 | 1 | 0 |
| 2 | 34 | 4 | 0 | 1 | 0 | 1 | 3 |
| 3 | 41 | 4 | 1 | 0 | 0 | 0 | 4 |
| 4 | 30 | 0 | 1 | 1 | 0 | 0 | 1 |

```
X = df.drop(columns = ['Jenis_KB'])
y = df['Jenis_KB']

print("X : ", X.shape)
print("y : ", y.shape)
```

```
X :  (254, 6)
y :  (254,)
```

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

```
print(f"x_train : {x_train.shape}")
print(f"y_train : {y_train.shape}")
print(f"x_test : {x_test.shape}")
print(f"y_test : {y_test.shape}")
```

```
x_train : (203, 6)
y_train : (203,)
x_test : (51, 6)
y_test : (51,)
```

Gambar 8. Preparation Data

Dari data di atas kita merubah data string menjadi integer agar proses *Preparation Data* menjadi lebih mudah.misalnya jenis_kb dirubah menjadi interger. Pil Kb menjadi 2, IUD di rubah menjadi 0, Suntuk KB menjadi angka 3 dan seterusnya.

### 3.6 Modeling dan Evaluasi

Dalam mode ini kita memakai Confusion Matrix. Confusion matrix ialah sebuah tabel yang kerap dipakai untuk menghitung performa dari mode kategorisasi di machine learning. Tabel memvisualisasikan lebih detil mengenai jumlah data yang dikelompokkan betul atau salah [11], [12].

```
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(x_train, y_train)

y_pred = knn.predict(x_test)
KNN_acc = accuracy_score(y_pred, y_test)

print(classification_report(y_test, y_pred))
print('Akurasi KNN : {:.2f}%'.format(KNN_acc*100))
```

```
              precision    recall  f1-score   support

           0       0.77      1.00      0.87        10
           1       0.33      0.29      0.31         7
           2       0.56      0.56      0.56         9
           3       0.88      1.00      0.93         7
           4       1.00      0.86      0.92        14
           5       1.00      0.75      0.86         4

    accuracy                           0.76        51
   macro avg       0.76      0.74      0.74        51
weighted avg       0.77      0.76      0.76        51

Akurasi KNN : 76.47%
```

Gambar 9. Confusion Matrix

Gambar 9 menunjukkan data yang di dapat adalah Precision adalah 77% ,recall 76% dan accuracy 76.47% yang berarti data ini tergolong data *Fair Classification*

### 3.7 Testing

*Testing* adalah langkah penting dalam proses data mining untuk memastikan bahwa model yang dihasilkan memiliki kemampuan generalisasi yang baik dan dapat memberikan prediksi yang akurat pada data baru yang belum pernah dilihat sebelumnya.


Gambar 11. Testing Data

Dari gambar di atas kita akan memasukan data baru untuk menguji akurasi data menggunakan *K-Nearest Neighbors (KNN).*kita masukan data baru yaitu usia, jumlah anak, Kesehatan reproduksi, preferensi_durasi,Riwayat_penyakit dan Aktifitas Seksual.Maka hasil prediksi dari data di atas bernilai 4 yang berarti bahwa pasien tersebut di rekomendasikan yaitu Tubektomi.

## 4. PENUTUP

Penelitian ini telah berhasil menunjukkan penerapan algoritma *K-Nearest Neighbors* (KNN) untuk menentukan jenis kontrasepsi (KB) yang sesuai berdasarkan karakteristik individu, seperti usia, jumlah anak, riwayat kesehatan, dan preferensi pribadi. Dengan menggunakan platform Google Colab , proses analisis data dilakukan secara efisien, mulai dari preprocessing hingga evaluasi model. Hasil penelitian menunjukkan bahwa algoritma KNN mampu memberikan rekomendasi KB dengan tingkat akurasi yang baik, serta dapat dijadikan sebagai solusi alternatif dalam mendukung pengambilan keputusan di bidang kesehatan reproduksi. Evaluasi performa model menggunakan confusion matrix menunjukkan bahwa metode ini memiliki potensi untuk menghasilkan prediksi yang lebih personal dan relevan dibandingkan pendekatan manual tradisional.Namun, penelitian ini juga memiliki beberapa keterbatasan, seperti sensitivitas terhadap noise dalam dataset dan kebutuhan normalisasi fitur untuk meningkatkan hasil prediksi. Oleh karena itu, penelitian lanjutan dapat difokuskan pada eksplorasi algoritma lain atau kombinasi beberapa metode machine learning untuk meningkatkan akurasi dan robustness model. Selain itu, penggunaan dataset yang lebih besar dan representatif dapat membantu meningkatkan generalisasi model. Diharapkan penelitian ini dapat menjadi dasar pengembangan sistem rekomendasi KB berbasis teknologi yang lebih cerdas, sehingga dapat mendukung peningkatan layanan kesehatan reproduksi secara efektif dan inklusif.Dengan inovasi ini, diharapkan teknologi machine learning dapat semakin dimanfaatkan untuk memberikan solusi praktis dalam berbagai aspek kehidupan, termasuk perencanaan keluarga dan kesehatan masyarakat.

## DAFTAR PUSTAKA

[1] V. Kantorová, M. C. Wheldon, P. Ueffing, and A. N. Z. Dasgupta, “Estimating progress towards meeting women’s contraceptive needs in 185 countries: A Bayesian hierarchical modelling study,” *PLoS Med*, vol. 17, no. 2, 2020, doi: 10.1371/JOURNAL.PMED.1003026.

[2] J. V Raj, J. V. J. Anton, and J. P. Durai Raj, “Detection of recovery of covid-19 cases using machine learning,” *Int J Curr Res Rev*, vol. 13, no. 6 special Issue, 2021, doi: 10.31782/IJCRR.2021.SP183.

[3] J. R. Gewirtz O’Brien, “Missed Opportunities to Provide Comprehensive Sexual and Reproductive Healthcare Among Hospitalized Adolescents,” *J. Adolesc. Heal.*, vol. 68, no. 2, 2021, doi: 10.1016/j.jadohealth.2020.12.029.

[4] A. Iyer, “Data centric nanocomposites design: Via mixed-variable Bayesian optimization,” *Mol Syst Des Eng*, vol. 5, no. 8, 2020, doi: 10.1039/d0me00079e.

[5] R. J. Alfirdausy and S. Bahri, “Implementasi Algoritma K-Nearest Neighbor untuk Klasifikasi Diagnosis Penyakit Alzheimer,” *Techno.Com*, vol. 22, no. 3, 2023, doi: 10.33633/tc.v22i3.8393.

[6] A. N. Nan and D. Juniati, “KLASIFIKASI JENIS JANGKRIK BERDASARKAN SUARA

MENGGUNAKAN DIMENSI FRAKTAL METODE HIGUCHI DAN K-NEAREST NEIGHBOR (KNN)," *MATHunesa J. Ilm. Mat.*, vol. 10, no. 1, 2022, doi: 10.26740/mathunesa.v10n1.p199-207.

[7] L. Dausu, "Kesetaraan Gender dalam Program Keluarga Berencana di Kecamatan Wabula Kabupaten Buton," *Kybernan J. Stud. Kepemerintahan*, vol. 3, no. 2, 2020, doi: 10.35326/kybernan.v3i2.817.

[8] T. R. Abdillah, "Analisis Komparasi Cycles X Render Dan Cycles Render Menggunakan Google Colab," *J. TIKA*, vol. 8, no. 1, 2023, doi: 10.51179/tika.v8i1.1937.

[9] M. A. Khalimi, "Perhitungan Confusion Matrix Multi-Class Clasification 3x3."

[10] A. Thariq, "Implementasi Market Basket Analysis Menggunakan Algoritma Apriori pada Data Penjualan Buku," *J. Kolaboratif Sains*, vol. 6, no. 3, 2023.

[11] N. Iriadi, "Penerapan Algoritma Klasifikasi Data Mining Dalam Penentuan Pemberian Pinjaman Koperasi," *Paradigma*, vol. XIV, no. 2, 2012.

[12] P. Irfansyah, Y. A. Purwanto, and S. H. Wijaya, "Machine Learning Model to Determine Dominant Features in Palm Kernel Cake Quality," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1359, no. 1, 2024, doi: 10.1088/1755-1315/1359/1/012029.