

---

## KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS

Naisah Marito Putry<sup>1\*</sup>, Betha Nurina Sari, M.Kom<sup>2</sup>

<sup>1,2</sup> Teknik Informatika, Universitas Singaperbangsa Karawang  
Indonesia

\* Corresponding Author. E-mail: naisah.putry18071@student.unsika.ac.id

### Abstrak

Diabetes melitus merupakan penyakit gangguan metabolik yang disebabkan oleh pankreas yang sudah tidak dapat memproduksi cukup insulin atau tubuh sudah tidak dapat menggunakan insulin secara efektif. Penelitian ini menggunakan dua algoritma yaitu Naïve Bayes dan KNN. Hal ini untuk membandingkan antara kedua algoritma tersebut yang memiliki tingkat akurasi yang terbaik. Selain itu, kedua algoritma tersebut digunakan untuk dapat menghasilkan informasi dari dataset diabetes melitus yang digunakan. Algoritma Naïve Bayes adalah salah satu algoritma yang digunakan untuk klasifikasi statistika yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Dan algoritma KNN adalah algoritma *supervised learning* yang digunakan untuk klasifikasi objek baru berdasarkan objek terdekatnya. Pada penelitian ini proses klasifikasi dilakukan dengan memasukkan data ke dalam *Jupyter Notebook* dan membuat rancangan prosesnya, kemudian data tersebut akan diolah dengan algoritma KNN dan Naïve Bayes. Dan tahapan penelitian yang digunakan dalam penelitian ini adalah KDD. Nilai akurasi yang dihasilkan berbeda, dan nilai akurasi algoritma Naïve Bayes lebih tinggi dibandingkan algoritma KNN.

Kata kunci : Klasifikasi; Diabetes Melitus; Algoritma Naïve Bayes; Algoritma KNN, KDD.

### Abstract

*Diabetes melitus is a metabolic disorder caused by the pancreas not being able to produce enough insulin or the body being unable to use insulin effectively. This study uses two algorithms, namely Naïve Bayes and KNN. This is to compare the two algorithms that have the best level of accuracy. In addition, both algorithms are used to generate information from the diabetes melitus dataset used. Naïve Bayes algorithm is one of the algorithms used for statistical classification that can be used to predict the probability of membership of a class. And the KNN algorithm is a supervised learning algorithm that is used to classify new objects based on their closest objects. In this study, the classification process is carried out by entering data into the Jupyter Notebook and designing the process, then the data will be processed using the KNN and Naïve Bayes algorithms. And the research stages used in this research are KDD. The resulting accuracy values are different, and the accuracy value of the Naïve Bayes algorithm is higher than the KNN algorithm.*

Keywords : Classification; Diabetes Melitus; Naïve Bayes Algorithm; KNN Algorithm; KDD.

## 1. Introduction

Diabetes melitus adalah sebuah penyakit metabolik yang disebabkan oleh karena ketidakmampuan tubuh dalam memanfaatkan insulin atau kekurangan hormon insulin, akibatnya kadar gula di dalam darah tidak dapat terkendali [1].

Penyakit diabetes kini menyerang manusia tanpa mengenal usia. Bahkan lebih dari 1,2 juta anak-anak dan remaja di dunia terkena penyakit diabetes [2]. Penyakit diabetes pun masih masuk ke daftar penyakit paling mematikan di dunia [3]. Pada tahun 2021, jumlah pengidap yang meninggal dikarenakan diabetes di dunia mencapai 6,7 juta orang [4].

Menurut *World Health Organization* jumlah penderita penyakit diabetes melitus terus saja bertambah banyak hingga empat kali lebih banyak dibandingkan tahun-tahun sebelumnya. Penyakit diabetes melitus membunuh satu juta orang untuk setiap tahunnya. Penyakit diabetes melitus dapat disembuhkan asalkan penderita sudah menyadari dan mengikuti pengobatan sejak awal sebelum penyakit tersebut bertambah parah. Oleh karena itu, perlunya informasi terhadap masyarakat mengenai faktor-faktor penyebab penyakit diabetes melitus [5].

Banyak penelitian yang telah dilakukan untuk mengidentifikasi penyakit diabetes melitus, yaitu dapat dilihat pada tabel 1.

**Tabel 1.** Penelitian Sebelumnya

Judul	Keterangan
Penerapan Metode <i>K-Nearest Neighbor</i> Untuk Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus	Penelitian ini menggunakan algoritma <i>K-Nearest Neighbor</i> untuk mengolah dataset penyakit diabetes melitus dan menghasilkan nilai akurasi sebesar 96% dengan nilai $k = 23$ [6].
Implementasi Algoritma <i>K-Nearest Neighbor</i> Untuk Memprediksi Pasien Terkena Penyakit Diabetes pada Puskesmas Menyampa Kabupaten Bulukumba	Penelitian ini menggunakan algoritma <i>K-Nearest Neighbor</i> dan memiliki tingkat akurasi sebesar 68,3% [7].
Penerapan Metode <i>Particle Swarm Optimization</i> Untuk Meningkatkan Kinerja Algoritma <i>K-Nearest Neighbor</i> Dalam Klasifikasi Penyakit Diabetes	Penelitian ini menggunakan algoritma KNN, dengan menggunakan nilai $k = 19$ dan hasil akurasi yang dihasilkan sebesar 77,213% [8].

Perancangan Sistem Penelitian ini Pakar Diagnosis menggunakan Penyakit Diabetes algoritma *Naïve Berbasis Web Bayes* untuk Menggunakan melakukan diagnosis Algoritma *Naïve* penyakit diabetes *Bayes* dan menghasilkan tingkat akurasi yang cukup tinggi [9].

Penerapan Penelitian ini Algoritma *Naïve* menggunakan *Bayes* Untuk algoritma *Naïve* Klasifikasi Penyakit *Bayes* yang memiliki Diabetes Mellitus nilai akurasi yang tinggi yaitu 90,20% [10].

Penerapan Penelitian ini juga Algoritma *Naïve* menggunakan *Bayes* Untuk algoritma *Naïve* Klasifikasi Penyakit *Bayes* dengan Diabetes Mellitus di menghasilkan nilai Rumah Sakit akurasi sebesar Aisyiah 90,20% [11].

Prediksi Pima Penelitian ini Indians Diabetes menyatakan bahwa Database Dengan pengujian presisi *Ensemble Adaboost* penggunaan algoritma *Naïve Bayes* lebih baik tanpa menggunakan *ensemble* baik *bagging* dan *adaboost* dengan

nilai akurasi sebesar 80,23% [12]. Analisis Penelitian ini Rekomendasi menggunakan 2 Penerima Beasiswa algoritma yaitu C4.5 Menggunakan dan KNN dimana Algoritma *K-Nearest Neighbor* dihasilkan bahwa nilai akurasi KNN dan Algoritma C4.5 lebih tinggi dibandingkan C4.5 yaitu 90,7% [13].

Perbandingan Penelitian ini Akurasi Algoritma melakukan C4.5 Dan *Naïve* perbandingan antara *Bayes* Untuk Deteksi algoritma *Naïve* Dini Gangguan *Bayes* dan C4.5, dan Autisme Pada Anak menghasilkan algoritma *Naïve Bayes* nilai akurasinya lebih tinggi yaitu sebesar 73,33% [14].

Berdasarkan penelitian sebelumnya yang telah dijabarkan pada tabel 1 dapat dilihat bahwa algoritma KNN dan *Naïve Bayes* memiliki keunggulan masing-masing. Ada penelitian yang sudah membandingkan kedua algoritma tersebut dengan algoritma C4.5, hasilnya menunjukkan bahwa nilai akurasi KNN dan *Naïve Bayes* lebih unggul.

Penelitian ini akan melakukan perbandingan antara algoritma KNN dan

*Naïve Bayes* dalam klasifikasi penyakit diabetes melitus. Penggunaan algoritma KNN ini dikarenakan KNN merupakan algoritma yang efektif untuk data yang besar, tahan terhadap data pelatihan yang *noise*, dan memiliki performa yang baik. Sedangkan untuk penggunaan algoritma *Naïve Bayes* dikarenakan cepat dan efisien terhadap ruang, selain itu hanya memerlukan sedikit data pelatihan untuk melakukan klasifikasi.

Belum ada penelitian sebelumnya yang membandingkan performa algoritma KNN dan *Naïve Bayes* dalam mengklasifikasikan penyakit diabetes melitus. Diharapkan dengan adanya penelitian ini dapat memberikan informasi yang bermanfaat bagi masyarakat mengenai penyakit diabetes melitus dan penelitian ini dapat menjadi acuan referensi serta pengembangan ilmu pengetahuan untuk penelitian selanjutnya.

## **2. Materials and Methods**

### **2.1 Materials.**

#### **A. Penyakit**

Penyakit yaitu suatu keadaan yang terjadi akibat adanya gangguan atau kegagalan mekanisme terhadap keseimbangan fungsi tubuh atau sistem tubuh sehingga tubuh tidak dalam keadaan yang normal. Penyakit merupakan suatu tekanan sehingga dapat menimbulkan gangguan pada fungsi sistem tubuh,

penyakit tidak hanya dilihat dari luar saja tetapi juga adanya ketidakaturan fungsi-fungsi dalam tubuh [15].

Adapun pengertian lain dari penyakit yaitu suatu kondisi buruk yang disebabkan oleh adanya mikro organisme berbahaya di dalam organ tubuh tertentu sehingga terjadi munculnya sel yang tidak sempurna di dalam tubuh dan terjadinya ketidakseimbangan kimiawi di dalam tubuh [16].

#### **B. Diabetes Melitus**

Penyakit diabetes melitus (kencing manis) adalah salah satu penyakit yang berbahaya dan banyak membunuh satu juta orang untuk setiap tahunnya. Menurut *World Health Organization* jumlah penderita terus saja bertambah banyak hingga empat kali lebih banyak dari pada tahun-tahun sebelumnya[5].

Penyebab terjadinya diabetes melitus ini adalah kadar gula darah di dalam tubuh yang tinggi. Diabetes melitus adalah sebuah penyakit metabolik yang disebabkan oleh karena ketidakmampuan tubuh dalam memanfaatkan insulin atau kekurangan hormon insulin, akibatnya kadar gula di dalam darah tidak dapat terkendali. Insulin merupakan hormon yang dihasilkan oleh sel beta yang ada di dalam pankreas, dan insulin akan memberikan sinyal ke sel-sel yang ada di tubuh agar sel-sel tersebut menyerap glukosa [17].

Diabetes melitus merupakan kelainan heterogen yang ditandai dengan tingginya kadar gula yang ada di dalam tubuh. Diabetes merupakan penyakit metabolik kronis, apabila tidak dilakukan perawatan dan pengobatan yang tepat dan cepat akan mengakibatkan kondisi yang sangat membahayakan pasien dan juga dapat menyebabkan komplikasi [18].

### C. Data

Data merupakan sekumpulan fakta yang didapatkan dari suatu pengukuran. Suatu pengambilan keputusan yang tepat adalah hasil dari penarikan kesimpulan yang berdasarkan pada data atau fakta yang akurat. Dapat dikatakan bahwa data adalah kenyataan yang menggambarkan suatu kejadian atau suatu bentuk yang masih mentah dan perlu diolah lebih lanjut agar dapat menghasilkan suatu informasi [19].

### D. Klasifikasi

Klasifikasi adalah suatu proses pembelajaran yang memiliki fungsi untuk menentukan setiap himpunan atribut dari sebuah objek atau dapat dikatakan juga klasifikasi merupakan proses pengelompokan suatu data. Pengelompokan adalah kegiatan penempatan suatu objek ke dalam kelas yang memiliki fungsi yang sama. Klasifikasi digunakan dalam mendeskripsikan sebuah dataset dimana setiap tipe data yaitu berbentuk nominal

atau biner. Dalam klasifikasi data *supervised* akan dibagi menjadi dua, yaitu *data training* dan *data testing*, dimana *data training* akan dianalisis dengan menggunakan algoritma klasifikasi. Berikut beberapa algoritma klasifikasi yaitu algoritma KNN, algoritma Naïve Bayes, algoritma C4.5, algoritma C5, dan sebagainya [20].

### E. Data Mining

Data mining adalah proses pengelolaan sebuah data yang besar sehingga dari data-data tersebut dapat memberikan informasi yang akurat dan dapat mempermudah dalam pemecahan masalah dan pengambilan keputusan. Adapun nama-nama lain dari *data mining* adalah *Pattern Analysis*, *Knowledge Extraction*, *Information Harvesting*, dan sebagainya. Terdapat enam pengelompokan dalam *data mining* yaitu [21]:

#### 1. Estimasi

Untuk kelompok estimasi hampir sama dengan klasifikasi, kecuali *variable target* dari estimasi lebih mengarah ke numerik daripada ke arah kategori. Estimasi membangun sebuah model dimana model tersebut didapatkan dari *record* lengkap yang menyediakan nilai yang diambil dari variabel target sebagai nilai prediksi.

#### 2. Deskripsi

Kelompok deskripsi adalah cara untuk menggambarkan pola yang terdapat di dalam data yang dimiliki.

### 3. Klasifikasi

Klasifikasi adalah suatu proses pembelajaran yang memiliki fungsi untuk menentukan setiap himpunan atribut dari sebuah objek atau dapat dikatakan juga klasifikasi merupakan proses pengelompokan suatu data.

### 4. Prediksi

Digunakan untuk memperkirakan nilai di masa depan atau masa mendatang dan juga menerka nilai yang belum diketahui.

### 5. Asosiasi

Asosiasi digunakan untuk menemukan atribut yang muncul dalam satu waktu.

### 6. Clustering

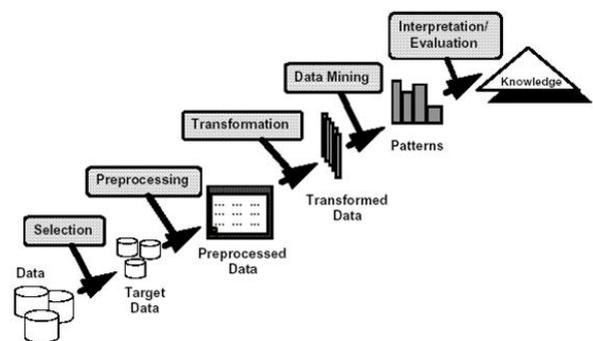
Clustering adalah pengamatan, pengelompokan *record*, membentuk kelas objek-objek yang mempunyai kemiripan.

Selain itu, data mining memiliki beberapa jenis data sebagai berikut :

1. *Semi supervised* adalah perpaduan data *unsupervised* dan *supervised*.
2. *Supervised learning* adalah jenis data yang sudah memiliki label atau kelas.
3. *Unsupervised learning* adalah jenis data yang tidak memiliki label atau kelas.

## F. Knowledge Discovery in Database

*Knowledge Discovery in Database* (KDD) adalah metode yang digunakan untuk mendapatkan pengetahuan dari sebuah basis data. Hasil akhir dari proses KDD yaitu untuk bahan pertimbangan dalam pengambilan keputusan. Pada tahapan KDD memiliki rangkaian proses seperti Gambar 1 [22].



Gambar 1. Tahap KDD

### 1. Data Selection

Tahap seleksi data (*data selection*) adalah tahapan yang memilah-milah data yang akan digunakan dalam proses data mining, kemudian data hasil seleksi akan dibuat terpisah dengan *database* operasional.

### 2. Data Preprocessing

Tahap *data preprocessing* meliputi proses pembersihan (*cleaning*) dengan memperbaiki kesalahan yang ada pada data, membuang duplikasi data, dan memeriksa data-data yang tidak konsisten. Selain itu, pada tahap ini data-data yang sudah ada akan dilakukan *enrichment* dengan

informasi atau data eksternal yang dibutuhkan.

### 3. *Transformation Data*

Transformasi data adalah proses transformasi atau perubahan bentuk data yang memiliki entitas yang belum jelas ke bentuk data yang siap atau valid untuk diproses di langkah selanjutnya (*data mining*).

### 4. *Data mining*

*Data mining* adalah proses yang mengelola data menjadi informasi dengan terdapat beragam metode dan teknik. *Data mining* adalah proses pengelolaan sebuah data yang besar sehingga dari data-data tersebut dapat memberikan informasi yang akurat dan dapat mempermudah dalam pemecahan masalah dan pengambilan keputusan [23].

### 5. *Knowledge Interpretation atau Evaluation*

Informasi yang diperoleh dari proses *data mining* diperiksa untuk menghindari adanya informasi yang bertentangan dengan fakta yang ada sebelumnya. Dan informasi akan ditampilkan dalam bentuk yang mudah dipahami.

## G. *Naïve Bayes*

*Naïve bayes* merupakan algoritma yang digunakan untuk klasifikasi statistika yang dapat digunakan untuk memprediksi

probabilitas keanggotaan suatu *class*. *Naïve bayes* sendiri sudah terbukti mempunyai kecepatan dan akurasi yang tinggi. Ketika digunakan untuk *database* dengan data yang besar. Untuk klasifikasi *naïve bayes* ini memiliki kemampuan klasifikasi yang serupa dengan *neural network* dan *decision tree*. Berikut rumus Teorema Bayes [20]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

## H. *K-Nearest Neighbor (KNN)*

Algoritma *K-Nearest Neighbor* merupakan salah satu metode klasifikasi *data mining*, KNN mengklasifikasikan sekumpulan data berdasarkan data pembelajaran diberi label. KNN termasuk ke dalam *supervised learning* yang digunakan untuk klasifikasi objek baru berdasarkan objek terdekatnya. Hasil *query instance* yang baru, akan diklasifikasikan berdasarkan yang paling banyak jumlahnya atau mayoritas dari kategori pada KNN, dapat diartikan juga kelas yang paling banyak muncul akan dijadikan sebagai kelas klasifikasi. KNN juga merupakan suatu pendekatan yang digunakan untuk menghitung nilai kedekatan antara kasus lama dengan kasus baru. Pengklasifikasian KNN tidak menggunakan model hanya berdasarkan pada memori. Algoritma KNN menggunakan klasifikasi ketetanggaan yang

digunakan sebagai nilai prediksi dari sampel uji yang baru [24].

Berikut langkah-langkah melakukan klasifikasi dengan menggunakan algoritma *K-Nearest Neighbor* (KNN) [25]:

1. Tentukan parameter *K* yang sesuai dengan data yang digunakan. Nilai *K* minimal bernilai 1 dan maksimalnya jumlah dari data latih.
2. Hitunglah jarak antara data uji dan data latih. Untuk menghitung jarak tersebut di dalam perhitungan algoritma KNN biasanya menggunakan perhitungan jarak *Euclidean* dengan rumus seperti persamaan 2.2.

$$Euclidean = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2)$$

Keterangan :

$p_i$  = data latih

$q_i$  = data uji

$i$  = data variabel

$n$  = dimensi data.

3. Kemudian jarak tersebut diurutkan dari yang terbesar ke yang terkecil.
4. Menentukan jarak terdekat sampai pada parameter *K*.
5. Pasangkan kelas yang sesuai.
6. Mencari jumlah kelas dari tetangga terdekat dan menetapkan kelas tersebut sebagai kelas data yang akan diklasifikasikan atau dievaluasi.

## 2.2. Methods

Objek penelitian yang dilakukan dalam penelitian ini adalah diagnosis penyakit diabetes melitus. Data yang digunakan yaitu berasal dari *website* Kaggle.com. Dataset berasal dari *National Institute of Diabetes and Digestive and Kidney Diseases* [26]. Total data yang digunakan adalah 200 data. Dimana variable yang digunakan *Outcome, Glucose, Blood Pressure, Insulin, BMI, Age, dan Diabetes Pedigree Function*. Dan *tools* yang digunakan dalam penelitian ini adalah *Jupyter Notebook*. Data akan diolah dengan menggunakan algoritma *K-Nearest Neighbor* dan *Naïve Bayes*. data tersebut akan dilakukan teknik data mining untuk mengklasifikasikan diagnosis penyakit diabetes melitus. Metodologi penelitian yang digunakan adalah *Knowledge Discovery in Database* (KDD).

### 1. Data Selection

Langkah pertama yang dilakukan sebelum mengumpulkan data yang akan digunakan dalam penelitian yaitu melakukan studi literatur yang bersumber dari jurnal, buku, dan sebagainya. Dan mencari data secara daring dari *website* Kaggle.com sesuai dengan yang akan digunakan dalam penelitian. Selain itu, menentukan atribut-atribut yang akan digunakan dalam penelitian ini. Atribut yang digunakan sebagai label dalam penelitian ini adalah *Outcome*.

## 2. Data Preprocessing

Pada tahap ini meliputi proses pembersihan data. Hanya melakukan penghapusan beberapa atribut (kolom) yang tidak digunakan dalam penelitian ini. Atribut yang tidak digunakan dalam penelitian ini adalah *Pregnancies* dan *Skin Thickness*. Tetapi pada penelitian ini data yang digunakan sudah lengkap dan tidak terdapat *missing value*.

## 3. Transformation Data

Melakukan perubahan pada atribut *Outcome* dari bentuk angka (0 & 1) menjadi *yes & no*.

**Tabel 2.** Data awal diabetes melitus

V1	V2	V3	V4	V5	V6	V7
0	89	76	37	31.200	0.192	23
1	88	30	99	55	0.496	26
0	118	58	94	33.300	0.261	23
1	117	88	145	34.500	0.403	40
0	136	74	204	37.400	0.3999	24

Sumber : Kaggle Diabetes Dataset

Keterangan :

- 1) V1 = *Outcome*
- 2) V2 = *Glucose*
- 3) V3 = *Blood Pressure*
- 4) V4 = *Insulin*
- 5) V5 = *BMI*
- 6) V6 = *Diabetes Pedigree Function*
- 7) V7 = *Age*.

Table 2 merupakan bentuk data awal dari diabetes melitus, kemudian dilakukan

perubahan pada bagian atribut *outcome* sehingga dihasilkan data seperti pada Tabel 3.

**Tabel 3.** Data diabetes melitus setelah transformasi

V1	V2	V3	V4	V5	V6	V7
no	89	76	37	31.200	0.192	23
yes	88	30	99	55	0.496	26
no	118	58	94	33.300	0.261	23
yes	117	88	145	34.500	0.403	40
no	136	74	204	37.400	0.3999	24

## 4. Data Mining

Pada tahap ini yaitu membuat model dengan menerapkan kedua algoritma klasifikasi yaitu algoritma KNN dan *Naïve Bayes*.

## 5. Evaluation

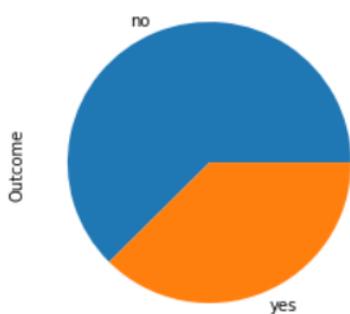
Pada tahap ini akan melakukan evaluasi terhadap model yang dibuat, dimana bentuk evaluasi yang dilakukan adalah menghitung nilai akurasi, *recall*, dan *precision* dari algoritma yang digunakan.

## 3. Results and Discussion

Penelitian ini akan mengklasifikasikan diagnosis penyakit diabetes melitus. Dimana yang menjadi data masukan atau *input* dalam penelitian ini adalah data diagnosis penyakit diabetes melitus dengan variable-variabel penentunya yaitu *Outcome*, *Glucose*, *Blood Pressure*, *Insulin*, *BMI*, *Age*, *Diabetes Pedigree Function*.

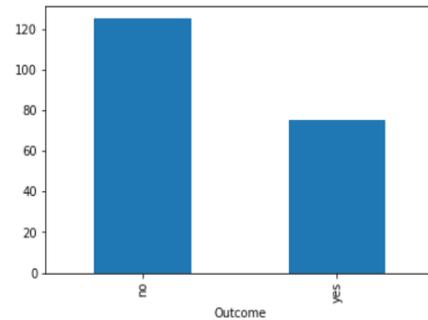
Keluarannya (*output*) memiliki 2 kategori yaitu *yes* dan *no*. Untuk keluaran *yes*, maka berdasarkan ciri-ciri (*diagnosis*) yang terdapat dalam data, bahwa orang tersebut menderita diabetes. Dan begitupun dengan sebaliknya yaitu bila keluaran *no*, maka berdasarkan ciri-ciri (*diagnosis*) yang terdapat dalam data, bahwa orang tersebut tidak menderita diabetes.

Setelah didapatkan data, maka data tersebut diolah dengan cara dibersihkan dari *outlier* dan *noise*. Kemudian setelah dilakukan *data cleaning*, maka yang dilakukan selanjutnya adalah pembuatan model (melakukan klasifikasi dengan algoritma KNN dan *Naïve Bayes*) dengan menggunakan *Jupyter Notebook* dan bahasa pemrograman Python. Berikut contoh visualisasi data dalam bentuk grafik lingkaran berdasarkan atribut *Outcome* (Gambar 2).



**Gambar 2.** Visualisasi data dalam grafik lingkaran

Berikut contoh visualisasi data dalam bentuk grafik batang berdasarkan atribut *Outcome* (Gambar 3).



**Gambar 3.** Visualisasi data dalam grafik batang

Kemudian pemodelan dengan menggunakan *software Jupyter Notebook* dengan bahasa pemrograman Python. Pemodelan pada penelitian ini adalah klasifikasi data dengan algoritma *Naïve Bayes* dan *K-Nearest Neighbor*. Nilai *k* untuk algoritma KNN adalah 6. Selain itu, data akan dievaluasi dengan menghitung nilai akurasi, *recall*, dan *precision* dari algoritma KNN dan *Naïve Bayes*. Berikut Tabel 4 hasil nilai akurasi dari KNN dan *Naïve Bayes*

**Tabel 4.** Nilai akurasi

Pembagian Data ( <i>Testing – Training</i> )	Akurasi <i>Naïve Bayes</i>	Akurasi KNN
10 – 90	80%	75%
20 – 80	78%	75%
30 – 70	80%	71%
40 – 60	74%	66%
50 – 50	78%	65%

Selanjutnya melakukan evaluasi dengan menghitung nilai *recall* dari masing-masing algoritma, hasil dapat dilihat pada Tabel 5.

**Tabel 5.** Nilai *recall*

Pembagian Data (Testing – Training)	<i>Recall</i> <i>Naïve</i> <i>Bayes</i>	<i>Recall</i> KNN
10 – 90	0.86	0.86
20 – 80	0.89	0.89
30 – 70	0.86	0.92
40 – 60	0.82	0.86
50 – 50	0.85	0.85

Dan evaluasi terakhir yang dilakukan adalah menghitung nilai *precision* dari masing-masing algoritma, hasil dapat dilihat pada Tabel 6.

**Tabel 6.** Nilai *precision*

Pembagian Data (Testing – Training)	<i>Precision</i> <i>Naïve</i> <i>Bayes</i>	<i>Precision</i> KNN
10 – 90	0.86	0.80
20 – 80	0.81	0.78
30 – 70	0.82	0.75
40 – 60	0.77	0.68
50 – 50	0.80	0.66

#### 4. Conclusion

Penelitian ini melakukan komparansi (perbandingan) antara dua algoritma yaitu KNN dan *Naïve Bayes* dalam mengklasifikasikan diagnosis penyakit diabetes melitus. Selain itu, penelitian ini menggunakan lima pembagian data yang dilakukan dapat dilihat bahwa nilai akurasi dari *Naïve Bayes* lebih tinggi dibandingkan KNN. Dimana nilai akurasi paling tinggi

yang didapatkan dari algoritma *Naïve Bayes* yaitu sebesar 80%. Sedangkan algoritma KNN nilai akurasi tertinggi yaitu sebesar 75%. Selain itu, diketahui bahwa nilai *recall* paling tinggi dihasilkan oleh algoritma KNN yaitu sebesar 0.92. Dan untuk nilai presisi lebih tinggi dihasilkan oleh algoritma *Naïve Bayes* yaitu 0.86. Dengan dilakukan penelitian ini dapat memberikan informasi bermanfaat mengenai penyakit diabetes melitus dan algoritma yang lebih unggul diantara algoritma KNN dan *Naïve Bayes* sehingga dapat menjadi acuan referensi serta pengembangan ilmu pengetahuan untuk penelitian selanjutnya.

#### References

- [1] A. A. Fajrin and A. Maulana, “Penerapan Data Mining untuk Analisis Pola Pembelian Konsumen dengan Algoritma FPGrowth pada Data Transaksi Penjualan Spare Part Motor,” *Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, pp. 27–36, 2018.
- [2] E. Elisa, “Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti,” *JOIN*, vol. 2, no. 1, pp. 36–41, 2017.
- [3] Asroni, B. M. Respati, and S. Riyadi, “Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta,” *J. Ilm. Fak. Tek. Univ. Muhammadiyah*, vol. 21, no. 2, pp. 158–165, 2018.
- [4] M. R. Ritonga, Solikhun, M. R.

- Lubis, and A. P. Windarto, "SISTEM PAKAR DIAGNOSA GEJALA AWAL PENYAKIT AKIBAT VIRUS PADA ANAK BERBASIS MOBILE DENGAN FORWARD CHAINING," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 2, no. 2, pp. 140–146, 2018.
- [5] A. Najib, D. Nurcahyono, and R. P. P. Setiawan, "KLASIFIKASI DIAGNOSA PENYAKIT DIABETES MELLITUS (DM) MENGGUNAKAN ALGORITMA C4.4," *J. Sains Terap. Teknol. Inf.*, vol. 11, no. 2, 2019.
- [6] U. I. Lestari, A. Y. Nadhiroh, and C. Novia, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 4, pp. 2071–2082, 2021.
- [7] M. S. Mustafa and W. I. Simpen, "Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba," *Pros. Semin. Ilm. Sist. Inf. DAN Teknol. Inf.*, vol. 8, no. 1, pp. 1–10, 2019.
- [8] Sutrisno, D. Susilowati, and Haerani, "Penerapan Particle Swarm Optimization Untuk Meningkatkan Kinerja Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Diabetes," 2021.
- [9] Y. B. Widodo, S. A. Anggraeini, and T. Sutabri, "Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes," *J. Teknol. Inform. dan Komput. MH. Thamrin*, vol. 7, no. 1, pp. 112–123, 2021.
- [10] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 4, no. 1, pp. 15–21, 2020.
- [11] A. Afif, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus di Rumah Sakit Aisyiah," *J. Ilmu Komput. dan Mat.*, vol. 1, no. 2, pp. 40–46, 2020.
- [12] Rousyati, A. N. Rais, E. Rahmawati, and R. F. Amir, "Prediksi Pima Indians Diabetes Database Dengan Ensemble Adaboost Dan Bagging," *Evolusi J. Sains dan Manaj.*, vol. 9, no. 2, pp. 36–42, 2021.
- [13] D. Noviana, Y. Susanti, and I. Susanto, "Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor dan Algoritma C4.5," 2019.
- [14] B. Sagara, D. Adidarma, and S. Budilaksono, "PERBANDINGAN AKURASI ALGORITMA C4.5 DAN NAÏVE BAYES UNTUK DETEKSI DINI GANGGUAN AUTISME PADA ANAK," *J. Komput. dan Inform.*, vol. 3, no. 1, pp. 119–128, 2019.
- [15] I. S. Permana and Y. Sumaryana, "SISTEM PAKAR UNTUK MENDIAGNOSA PENYAKIT KULIT DENGAN METODE FORWARD CHAINING," *Jumantaka*, vol. 1, no. 1, pp. 361–370, 2018.
- [16] M. R. Ritonga, M. R. Lubis, and A. P. Windarto, "SISTEM PAKAR DIAGNOSA GEJALA AWAL PENYAKIT AKIBAT VIRUS PADA ANAK BERBASIS MOBILE DENGAN FORWARD CHAINING," *J. Nas. Inform. dan Teknol. Jar.*, vol. 2, no. 2, pp. 140–

- 146, 2018.
- [17] R. A. Nugroho, Tarno, and A. Prahutama, "KLASIFIKASI PASIEN DIABETES MELLITUS MENGGUNAKAN METODE SMOOTH SUPPORT VECTOR MACHINE (SSVM)," *J. Gaussian*, vol. 6, no. 3, pp. 439–448, 2017.
- [18] F. Aris and Benyamin, "Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi," *J. Sist. Komput. dan Sist. Inf.*, vol. 1, no. 1, pp. 1–6, 2019.
- [19] Nawassyarif, M. Julkarnain, and K. R. Ananda, "Sistem Informasi Pengolahan Data Ternak Unit Pelaksana Teknis Produksi dan Kesehatan Hewan Berbasis Web," *J. JINTEKS*, vol. 2, no. 1, pp. 32–39, 2020.
- [20] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode *Naive Bayes*," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
- [21] A. A. Fajrin and A. Maulana, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma FPGrowth Pada Data Transaksi Penjualan Spare Part Motor," *Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, pp. 27–36, 2018.
- [22] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [23] Asroni, B. M. Respati, and S. Riyadi, "Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta," *J. Ilm. Fak. Tek. Univ. Muhammadiyah*, vol. 21, no. 2, pp. 158–165, 2018.
- [24] Yahya and W. P. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020.
- [25] A. M. S. I. Dewi and I. B. G. Dwidasmara, "Implementation Of The K-Nearest Neighbor (KNN) Algorithm For Classification Of Obesity Levels," *J. Elektron. Ilmu Komput. Udayana*, vol. 9, no. 2, pp. 277–284, 2020.
- [26] M. Akturk, "Diabetes Dataset," *Kaggle.com*, 2020. <https://www.kaggle.com/mathchi/diabetes-data-set>.