

ANALISIS PREDIKSI PENYAKIT DIABETES PADA WANITA MENGGUNAKAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR

Julianto Lemantara^{1*}, Titik Lusiani²

^{1,2} Universitas Dinamika; Jalan Raya Kedung Baruk No. 98, Surabaya; Telp: 031-8721831.

Received: 12 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Naïve Bayes, K-Nearest Neighbor, Diabetes, WEKA, Knowledge Discovery Databases.

Correspondent Email:

julianto@dinamika.ac.id

Abstrak. Diabetes Melitus merupakan salah satu isu kesehatan yang populer di Indonesia karena prevalensi penyakit ini meningkat setiap tahun. Berbagai penelitian menyebutkan penyakit berbahaya ini lebih rentan terjadi pada jenis kelamin perempuan. Untuk itu, penelitian ini ingin melakukan prediksi/klasifikasi penyakit diabetes pada wanita sebagai upaya pencegahan. Penelitian ini membandingkan metode Naïve Bayes dan K-Nearest Neighbor (KNN) untuk prediksi penyakit diabetes dengan dataset berjumlah 768, terdiri atas 500 data training dan 268 data testing. Naïve Bayes digunakan karena bisa menghasilkan akurasi yang tinggi dengan jumlah data pelatihan yang sedikit. Sementara itu, K-Nearest Neighbor digunakan karena nonlinear dan nonparametrik, serta mudah diterapkan. Kinerja kedua metode dibandingkan agar metode yang lebih baik dalam hal klasifikasi penyakit dapat ditemukan. Hasil penelitian menunjukkan metode Naïve Bayes memiliki tingkat akurasi yang lebih unggul yaitu 78,3582%, sedangkan metode K-Nearest Neighbor memiliki tingkat akurasi 0,3731% dibawah metode dengan Naïve Bayes yaitu 77,9851%.

Abstract. Diabetes Mellitus is a popular health issue in Indonesia because the prevalence of this disease increases every year. Various studies found that this dangerous disease is more susceptible to women. For this reason, this research wants to predict/classify diabetes in women as a prevention effort. This research compared the Naïve Bayes and K-Nearest Neighbor (KNN) methods for predicting diabetes with a dataset of 768, consisting of 500 training data and 268 testing data. Naïve Bayes was used because it can produce high accuracy with less training data. Meanwhile, K-Nearest Neighbor was used because it is nonlinear, nonparametric, and easy to implement. The performances of the two methods were compared, so that a better method for carrying out classification was identified. The research results showed that the Naïve Bayes method has a superior accuracy rate, namely 78.3582%, while the K-Nearest Neighbor method has an accuracy rate of 0.3731%, below the Naïve Bayes method, namely 77.9851%.

1. PENDAHULUAN

Diabetes Melitus merupakan salah satu isu kesehatan yang populer di negara Indonesia dikarenakan jumlah keseluruhan (prevalensi) penyakit diabetes meningkat setiap tahun [1]–[3]. Diabetes Melitus yaitu penyakit yang timbul karena adanya hiperglikemia. Penyakit ini disebabkan oleh gangguan sekresi insulin,

kerja insulin, ataupun kedua-duanya [4], [5]. Hiperglikemia yang kronis pada penyakit diabetes ini akan menyebabkan banyak kerusakan pada organ tubuh manusia misalnya: mata, ginjal, jantung, pembuluh darah, dan saraf [4], [6].

Diabetes melitus ditandai dengan adanya kelebihan batas normal pada kadar gula darah,

yaitu ≥ 200 mg/dl dan pada saat menjalankan puasa, yaitu ≥ 126 mg/dl [7], [8]. Pengidap kasus penyakit diabetes ini seringkali tidak menyadari hal tersebut sehingga diabetes melitus disebut silent killer [7], [9]. Banyak orang tidak menyadari bahwa dirinya telah terkena Diabetes Melitus pada awalnya. Pada tahun 2015, International Diabetes Federation (IDF) menyatakan bahwa dari prediksi penderita diabetes berjumlah 415 juta di dunia dengan usia antara 20 dan 79 tahun, hampir 50% yaitu 193 juta orang tidak mengetahui jika terkena penyakit Diabetes Melitus. Terdapat 318 juta orang dewasa diperkirakan sebenarnya telah mengidap masalah prediabetes atau toleransi gula [10].

Diabetes melitus dibagi ke dalam 2 jenis tipe yakni diabetes melitus jenis tipe 1, yang disebabkan sistem kekebalan tubuh yang merusak pankreas dalam memproduksi insulin, sedangkan diabetes jenis tipe 2 disebabkan karena sel-sel tubuh yang kurang sensitif terhadap insulin [11]. Pada diabetes tipe 1, autoimun akan menghancurkan insulin yang diproduksi sel-sel beta pankreas pada keadaan normal [12]. Sedangkan jenis tipe 2 disebabkan karena adanya kelemahan daya sensitivitas pada insulin (resisten insulin) atau kelemahan dalam pembentukan jumlah insulin [6], [13]. Aspek utamanya sebagai akibat dari obesitas atau kegemukan dan pola hidup yang kurang sehat [13], [14]. Faktor ini dapat diatasi dengan teratur olah raga dan pola diet sehat [15].

Menurut IDF, jumlah keseluruhan kasus diabetes di seluruh dunia adalah 1,9% sehingga wajar penyakit ini menjadi penyebab kematian urutan ketujuh di dunia [7], [15]. Di dunia kasus diabetes melitus mencapai 382 juta jiwa dengan rata-rata kasus diabetes melitus tipe 2. Rata-rata kasus diabetes tipe ini sangat besar yakni 85%-90% [7], bahkan ada yang menyebutkan jumlah penderitanya hingga mencapai 90%-95% [11], [15], [16].

Jumlah penderita Diabetes Melitus melonjak tajam. Fakta menunjukkan setiap 6 detik terdapat 1 orang meninggal akibat penyakit Diabetes Melitus. Di Indonesia, terdapat 4,5 juta penderita Diabetes Melitus di tahun 1995. Hal tersebut merupakan jumlah terbanyak ketujuh di seluruh dunia. Saat ini, angka tersebut meningkat hingga 10 juta dan diprediksi menjadi 16,2 juta di tahun 2040

sehingga Indonesia menjadi urutan keenam di seluruh dunia [17].

Hasil penelitian menunjukkan bahwa 59% pengidap diabetes melitus adalah perempuan [18]. Riset lainnya menunjukkan hasil yang senada bahwa jenis kelamin perempuan lebih rentan terkena penyakit diabetes dibandingkan jenis kelamin laki-laki dikarenakan faktor hormon dan metabolisme tubuh [19]. Sebuah riset juga menunjukkan 60,4% penderita diabetes berjenis kelamin perempuan dikarenakan perempuan mempunyai tingkat probabilitas peningkatan indeks massa tubuh yang lebih besar secara fisik [20]. Salah satu penelitian lainnya yang memperkuat bahwa pasien diabetes lebih banyak berjenis kelamin perempuan adalah penelitian dari Susilawati dan Rahmawati yang menyebutkan bahwa perempuan memiliki risiko 1,22 kali terkena diabetes dibandingkan dengan laki-laki [16].

Karena banyaknya wanita yang rentan terkena penyakit diabetes yang sangat berbahaya, maka penelitian ini bertujuan untuk memprediksi perempuan yang positif atau negatif diabetes. Data mining mempunyai solusi untuk melakukan prediksi perempuan yang positif atau negatif diabetes berdasarkan pola data-data sebelumnya. Data-data sebelumnya diolah menggunakan data mining dengan metode klasifikasi. Sebenarnya ada berbagai metode yang bisa diterapkan untuk pengklasifikasian dokumen. Namun, penelitian ini menerapkan metode Naïve Bayes dan K-Nearest Neighbor untuk melakukan prediksi perempuan positif atau negatif diabetes.

Metode Naive Bayes digunakan sebab dapat memberikan hasil akurasi yang tinggi dengan jumlah data training yang sedikit [21]–[23]. Metode ini diperkenalkan pertama kali oleh Thomas Bayes yaitu seorang ilmuwan Inggris. Naïve Bayes berhasil menjawab permasalahan-permasalahan nyata terkait statistik dan probabilitas. Pada penelitian ini, Naïve Bayes akan digunakan untuk menjawab persoalan penyakit Diabetes Melitus, terutama untuk kaum wanita.

Sementara itu, metode K-Nearest Neighbor dipilih karena sangat nonlinear, K-Nearest Neighbor adalah pembelajaran mesin yang bersifat nonparametrik pada algoritma (model) [24], [25]. Algoritma K-Nearest Neighbor merupakan algoritma yang mudah dimengerti dan juga mudah diterapkan [26].

Adapun penelitian terdahulu yang melakukan komparasi antara metode Naïve Bayes dan K-Nearest Neighbor adalah penelitian yang dilakukan untuk melakukan klasifikasi artikel jurnal berbahasa Indonesia [27]. Penelitian tersebut menggunakan 40 data jurnal, yang dipublikasikan dalam 2 tahun terakhir. Setiap jurnal diambil sebanyak 10 artikel. Hasil penelitian tersebut menemukan kinerja Naïve Bayes lebih baik dibandingkan dengan K-Nearest Neighbor setelah dilakukan percobaan pada kedua metode. Naïve Bayes menghasilkan klasifikasi sebanyak 28 dokumen dengan tepat (nilai akurasi 70%), sedangkan K-Nearest Neighbor menghasilkan klasifikasi sebanyak 16 dokumen dengan tepat (nilai akurasi 40%).

Penelitian sejenis lainnya yang juga telah melakukan komparasi antara metode Naïve Bayes dan K-Nearest Neighbor adalah penelitian yang dilakukan untuk melakukan klasifikasi waktu tunggu alumni dalam mendapatkan pekerjaan [28]. Hasil penelitian tersebut menunjukkan tingkat akurasi Naïve Bayes sebesar 87,50%, sedikit lebih unggul dibandingkan tingkat akurasi K-Nearest Neighbor sebesar 83,75%. Penelitian dengan perbandingan kedua metode ini juga diterapkan pula untuk melakukan klasifikasi status gizi balita di Puskesmas Muara Jawa [29]. Hasil penelitian ini menunjukkan kinerja K-Nearest Neighbor lebih unggul daripada Naïve Bayes dalam klasifikasi status gizi.

Perbedaan penelitian ini dengan beberapa penelitian terdahulu yaitu penelitian ini menggunakan Naïve Bayes dan K-Nearest Neighbor dengan tujuan untuk klasifikasi penyakit diabetes pada wanita. Penelitian lain yang sangat dekat, baik dari segi metode maupun segi tujuan penelitian yaitu penelitian yang dilakukan oleh Putry dan Sari [30]. Hanya saja penelitian Putry dan Sari ini menggunakan 6 atribut independent (glucose, insulin, blood pressure, diabetes pedigree function, BMI, dan age) serta 1 atribut dependent yaitu kelas diabetes. Sementara pada penelitian ini, menggunakan tambahan 2 atribut independent untuk membuat klasifikasi diabetes lebih komprehensif. Dua atribut yang dimaksud adalah triceps skin fold thickness dan number of times pregnant. Pada penelitian ini melibatkan atribut jumlah kehamilan dikarenakan

penelitian ini fokus untuk klasifikasi penyakit diabetes pada wanita saja.

2. TINJAUAN PUSTAKA

Data Mining

Data mining adalah proses analisis data berskala besar atau bersifat kompleks, yang bertujuan mendapatkan pola data yang penting dan seringkali keberadaannya tak disadari [31]. Data mining mempunyai beberapa istilah lain, yaitu pattern recognition/ knowledge discovery. Istilah pattern recognition (pengenalan pola) tepat untuk menggantikan data mining karena pengetahuan yang ingin diperoleh berbentuk pola-pola yang masih perlu digali dari data berskala besar. Semenatar itu, istilah knowledge discovery (penemuan pengetahuan) juga tepat digunakan untuk pengganti data mining sebab tujuan utama data mining yaitu memperoleh pengetahuan tersembunyi dari data berskala besar. Pada penelitian ini menggunakan istilah data mining dikarenakan istilah ini lebih populer dibandingkan dengan yang lainnya.

Naïve Bayes

Naïve Bayes adalah metode klasifikasi data dengan menggunakan teknik probabilistik yang sederhana. Metode ini menghitung sekumpulan peluang dengan melakukan penjumlahan frekuensi serta kombinasi nilai dari dataset [32]. Sebelum melakukan klasifikasi data, metode ini memerlukan data training. Metode ini mengasumsikan semua atribut pada masing-masing kategori tidak mempunyai ketergantungan satu dengan yang lain [33].

Pada saat proses training, data training telah ditentukan kelasnya, lalu diproses dan menghasilkan nilai probabilitas. Persamaan Teori Bayes dapat dilihat pada Persamaan 1 [34].

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

Keterangan:

$P(C|X)$ = Probabilitas hipotesis C berdasar kondisi X (probabilitas posterior)

C = Hipotesis data X merupakan suatu kelas spesifik

X = Data dengan kelas yang belum diketahui

$P(X|C)$ = Probabilitas X berdasarkan kondisi pada hipotesis C

$P(C)$ = Probabilitas hipotesis C (probabilitas prior)

$P(X)$ = Probabilitas X

Laplace Correction, atau sering disebut Laplacian Estimator/Additive Smoothing, biasa digunakan untuk menghindari nilai probabilitas nol (0) pada perhitungan metode Naïve Bayes disebabkan tidak adanya data kategori tertentu di kelasnya [23]. Untuk menerapkan teknik ini, hanya perlu menambah nilai 1 pada data untuk tiap-tiap kategori saat terdapat kategori mempunyai nilai 0. Laplace Correction dapat ditunjukkan pada Persamaan 2.

$$P(C|X) = \frac{P(X|C) \cdot P(C) + 1}{P(X) + N} \quad (2)$$

Keterangan:

$P(C|X)$ = Probabilitas hipotesis C berdasar kondisi X (probabilitas posterior)

C = Hipotesis data X merupakan suatu kelas spesifik

X : Data dengan kelas yang belum diketahui

$P(X|C)$ = Probabilitas X berdasarkan kondisi pada hipotesis C

$P(C)$ = Probabilitas hipotesis C (probabilitas prior)

$P(X)$ = Probabilitas X

N = banyaknya kelas (kategori)

K-Nearest Neighbor

Menemukan jarak antara dua titik yaitu titik training dan titik testing adalah prinsip kerja dari K-Nearest Neighbor [35]. Setelahnya evaluasi dilaksanakan dengan k tetangga terdekat pada data training. Persamaan jarak pada penelitian ini menggunakan rumus Euclidean distance, seperti yang tampak pada Persamaan 3.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Dimana :

d = jarak antara titik pada data training dan titik data testing

i = nilai atribut

n = dimensi atribut.

x_i = data training

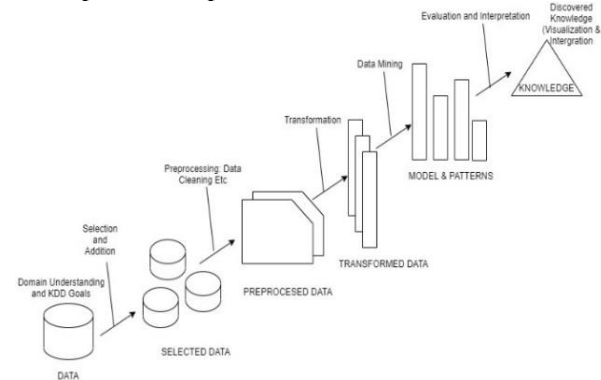
y_i = data testing

Penelitian ini menggunakan software WEKA untuk melakukan analisis dan

pengujian data. WEKA yaitu aplikasi data mining yang bersifat open-source. WEKA pertama kali dikembangkan dengan bahasa Java di Selandia Baru oleh Universitas Waikato.

3. METODE PENELITIAN

Knowledge Discovery Databases (KDD) adalah model proses data mining yang dipilih untuk digunakan pada penelitian ini, yang terdiri atas tahapan memahami domain dan tujuan, pemilihan dataset dan penentuan target, pemrosesan awal dan tranformasi data, penerapan metode data mining, serta evaluasi dan interpretasi. Lebih detailnya, tahapan KDD ini dapat dilihat pada Gambar 1.



Gambar 1 Tahapan KDD [36]

Pada penelitian ini menggunakan 768 dataset yang kemudian dibagi ke dalam 2 kategori yaitu:

1. 500 sebagai data training
2. 268 sebagai data testing

Memahami Domain Dan Tujuan

Penelitian ini bertujuan melakukan prediksi/klasifikasi wanita terkena diabetes atau tidak. Berdasarkan dataset diabetes yang diperoleh dari website Kaggle.com dengan jumlah 768 dataset menunjukkan bahwa data perempuan yang positif atau negatif diabetes sudah cukup lengkap dan representatif. Oleh karena itu, tujuan penelitian ini dapat dicapai dengan memanfaatkan data yang telah diperoleh.

Pemilihan Data Set dan Penentuan Target

Dataset yang digunakan dalam tahapan Knowledge Discovery Databases (KDD) adalah data wanita positif atau negatif terkena diabetes. Pemilihan data diabetes ini disebabkan karena mempunyai nilai kelengkapan data yang cukup

bagus. Data mentah yang didapat berjumlah 768, dengan masing-masing atribut yaitu:

1. Pregnancies: jumlah kehamilan
2. Glucose: tingkat gula darah, dalam mg/dL
3. Diastolic: tekanan darah diastolik, dalam mm Hg
4. Triceps: ketebalan lipatan kulit di trisep (mm), sebagai ukuran persentase lemak tubuh
5. Insulin: kadar insulin, dalam $\mu\text{U/mL}$
6. Body Mass Indeks (BMI): nilai terhitung yang diperoleh dari berat dan tinggi badan seseorang (berat dalam kilogram dibagi dibagi kuadrat tinggi badan dalam meter)
7. Diabetes pedigree function (DPF): ukuran memperkirakan keturunan diabetes berdasarkan riwayat keluarga.
8. Age: usia, dalam satuan tahun (years)
9. Kelas diabetes (terkena diabetes/tidak)

Pemrosesan Awal dan Transformasi Data

Setelah melakukan pengecekan terhadap setiap atribut yang dipilih, maka selanjutnya data dipastikan tidak boleh ada yang kosong. Pada penelitian ini, data tidak dilakukan proses pengelompokan atau pembuatan interval data karena data yang ada sudah beragam dan representatif untuk melakukan fungsi klasifikasi. Selain itu, semua data juga sudah seragam yaitu bertipe numerik. Adapun range/rentang data pada setiap atribut dapat dilihat pada Tabel 1.

Tabel 1. Rentang Nilai Setiap Atribut

No	Atribut	Rentang Nilai
1	Pregnancies	0 – 17
2	Glucose	0 – 199
3	Diastolic	0 – 122
4	Triceps	0 – 99
5	Insulin	0 – 846
6	BMI	0 – 67,1
7	DPF	0,08 – 2,42
8	Age	21 – 81
9	Diabetes Class	0 (tidak diabetes) 1 (diabetes)

Penerapan Metode Data Mining

Langkah pada saat pemilihan proses metode data mining. Ada beberapa metode dalam data mining yaitu klasifikasi, klustering, asosiasi, regresi dan lain-lain, objek sudah dapat dikategorikan saat data diperoleh dengan tujuan

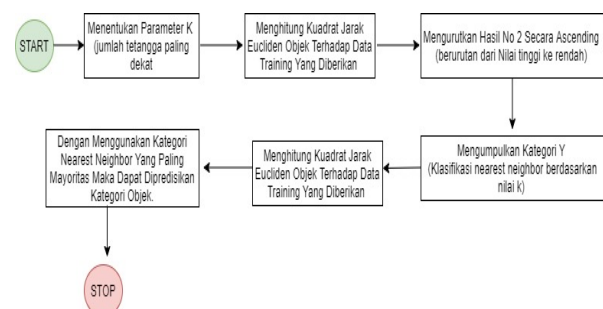
penemuan untuk model/pola baru dengan penghubungan antara data baru dan data yang sudah ada. Metode klasifikasi dipilih untuk membantu pengklasifikasian data perempuan yang positif Diabetes atau negative Diabetes.

Penelitian ini membandingkan algoritma Naïve Bayes dan K-Nearest Neighbor untuk memprediksi atau mengklasifikasikan penyakit diabetes yang kemudian hasil dari klasifikasi dapat digunakan untuk memprediksi apakah perempuan dalam keadaan tertentu positif atau negatif terkena penyakit Diabetes.

Pada tahap ini, sebelumnya data diolah terlebih dahulu sehingga hasilnya dapat digunakan untuk mengimplementasikan metode Naïve Bayes dan K-Nearest Neighbor. Tahapan Naïve Bayes dapat dilihat pada Gambar 2 dan tahapan K-Nearest Neighbor dapat dilihat pada Gambar 3.



Gambar 2 Tahapan Naïve Bayes



Gambar 3 Tahapan K-Nearset Neighbor

Evaluasi dan Interpretasi

Pada tahap evaluasi dan interpretasi ini, akan digunakan untuk mengevaluasi serta menginterpretasi hasil dari klasifikasi wanita yang positif atau negatif diabetes. Pada tahap ini

akan dievaluasi hasil kinerja klasifikasi antara metode Naïve Bayes dan K-Nearest Neighbor. Kinerja klasifikasi akan diukur berdasarkan tingkat akurasi, presisi, dan recall dengan menggunakan Confusion Matrix.

4. HASIL DAN PEMBAHASAN

Setelah melakukan tahap analisis dan pengujian data, maka dapat diketahui tingkat akurasi dari proses klasifikasi dengan metode Naïve Bayes dan K-Nearest Neighbor.

Berdasarkan 500 data training dan 268 data testing yang telah dianalisis dengan kedua metode, ditemukan hasil bahwa Naïve Bayes mampu melakukan klasifikasi dengan tepat untuk 210 data testing, sedangkan 58 data testing diklasifikasikan dengan tidak tepat sehingga tingkat akurasi dari metode Naïve Bayes ini adalah 78,3582%. Hasil pengolahan data dengan metode Naïve Bayes dapat terlihat pada Tabel 2. Untuk hasil akurasi dan pengukuran lain yang lebih detail dapat terlihat pada Tabel 3. Tabel 3 menunjukkan bahwa pembulatan nilai akurasi adalah 78,4%, nilai presisi adalah 77,9%, dan nilai recall adalah 78,4%.

Tabel 2. Summary Naïve Bayes

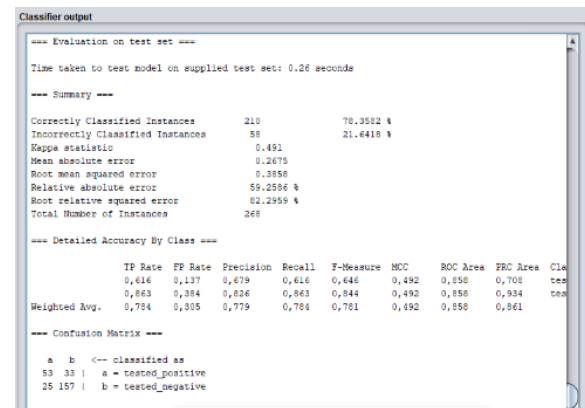
Correctly Classified Instances	210	78,3582%
Incorrectly Classified Instances	58	21,6418%
Kappa statistic	0,340972	
Mean absolute error	1,857639	
Root mean squared error	2,679167	
Relative absolute error	592586%	
Root relative squared error	822959%	
Total Number of Instances	268	

Tabel 3. Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall
	0,616	0,137	0,679	0,616
	0,863	0,384	0,826	0,863
Weighted Avg.	0,784	0,305	0,779	0,784

Sebagai bukti otentik bahwa hasil pengolahan data telah dilakukan dengan

Software Weka, maka diberikan tampilan output seperti yang tampak pada Gambar 4.



Gambar 4 Hasil Weka Naïve Bayes

Setelah pengolahan data dilakukan dengan metode Naïve Bayes, berikutnya penelitian ini melakukan pengolahan data dengan metode K-Nearest Neighbor, lalu membandingkan hasil prediksi/klasifikasi dari kedua metode tersebut. Untuk metode K-Nearest Neighbor ini sendiri berhasil melakukan klasifikasi 209 data dengan tepat, sedangkan 59 data lainnya diklasifikasikan dengan tidak tepat. Oleh karena itu, tingkat akurasi terbaik dari metode K-Nearest Neighbor dalam prediksi wanita yang terkena diabetes yaitu sebesar 77,9851%. Nilai K yang digunakan pada penelitian ini adalah 5 karena menghasilkan tingkat akurasi yang tertinggi dibandingkan nilai K yang lain yaitu 3, 7, dan 9.

Tingkat akurasi dari K-Nearest Neighbor ini hanya kalah tipis yaitu 0,3731% saja dari metode Naïve Bayes. Perbedaan akurasi dalam studi kasus ini bisa dikatakan tidak signifikan. Untuk lebih jelasnya, ringkasan hasil metode K-Nearest Neighbor (KNN) ini telah tersaji pada Tabel 4. Untuk hasil yang lebih detail yaitu tingkat akurasi, presisi, dan recall telah tersaji pada Tabel 5. Tabel 5 menunjukkan bahwa hasil pembulatan tingkat akurasi dari metode ini adalah 78%, tingkat presisi sebesar 77,3% dan tingkat recall sebesar 78%.

Tabel 4. Summary KNN

Correctly Classified Instances	209	77,9851%
Incorrectly Classified Instances	59	22,0149%
Kappa statistic	3,267361	

Mean absolute error	2,141667
Root mean squared error	2,820139
Relative absolute error	683007%
Root relative squared error	866126%
Total Number of Instances	268

Tabel 5. Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall
	0,57	0,121	0,69	0,57
	0,879	0,43	0,812	0,879
Weighted Avg.	0,78	0,331	0,773	0,78

Agar penelitian ini lebih meyakinkan secara ilmiah, tampilan output dari software Weka terkait pengolahan data dengan metode K-Nearest Neighbor ini diberikan pada Gambar 5.

```

Classifier output
--- Evaluation on test set ---
Time taken to test model on supplied test set: 0.73 seconds

--- Summary ---
Correctly Classified Instances      209      77.9851 %
Incorrectly Classified Instances    89       32.0149 %
Kappa statistic                    0.4705
Mean absolute error                 0.3084
Root mean squared error             0.4061
Relative absolute error             68.3007 %
Root relative squared error         86.6126 %
Total Number of Instances          268

--- Detailed Accuracy By Class ---

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,570	0,121	0,690	0,570	0,624	0,475	0,502	0,596	test
	0,879	0,430	0,812	0,879	0,844	0,475	0,502	0,565	test
Weighted Avg.	0,780	0,331	0,773	0,780	0,774	0,475	0,502	0,779	test

```

--- Confusion Matrix ---
a b <-- classified as
49 37 | a = tested_positive
22 160 | b = tested_negative

```

Gambar 5 Hasil Weka KNN

Setelah melakukan pengujian data, WEKA menampilkan confusion matrix. Dari confusion matrix inilah, nilai akurasi, presisi, dan recall dapat dihitung secara tepat. Rekap hasil pengolahan klasifikasi data dengan kedua metode dapat dilihat di tabel 6.

Tabel 6. Evaluasi Hasil Pengujian

Metode	Confusion Matrix		Accur ation	Precisi on	Recall
	a	b			
Naïve Bayes	53	33	78,4%	77,9%	78,4%
KNN	49	37	78%	77,3%	78%

Metode	Confusion Matrix	Accur ation	Precisi on	Recall
	22	160		

Keterangan :

a = tested positive

b = tested negative

Merujuk dari tabel 6, hasil penelitian ini menunjukkan bahwa metode Naïve Bayes dan K-Nearest Neighbor memiliki akurasi yang baik dalam melakukan klasifikasi/prediksi wanita yang terkena diabetes atau tidak. Kedua metode ini juga memiliki tingkat akurasi, presisi, dan recall yang tidak berbeda secara signifikan, meskipun metode Naïve Bayes sedikit lebih unggul dalam kasus penelitian ini. Hasil penelitian ini senada dengan penelitian [30] yang sama-sama bertujuan untuk melakukan klasifikasi penyakit diabetes. Pada penelitian [30] juga menemukan bahwa tingkat akurasi, presisi, recall dari metode Naïve Bayes lebih unggul dibandingkan dengan metode K-Nearest Neighbor, meskipun tidak selisih jauh. Tingkat akurasi yang terbaik dari metode Naïve Bayes adalah 80%, sedangkan tingkat akurasi yang terbaik dari K-Nearest Neighbor adalah 75%.

5. KESIMPULAN

Setelah tahapan penelitian telah dilalui secara keseluruhan, maka dapat diperoleh kesimpulan berikut ini:

- Baik Naïve Bayes maupun K-Nearest Neighbor sama-sama dapat digunakan untuk memprediksi wanita yang positif atau negatif diabetes karena memiliki akurasi yang baik.
- Metode Naïve Bayes memiliki nilai akurasi sebesar 78.3582% dan sedikit lebih tinggi dari metode K-Nearest Neighbor dengan nilai akurasi sebesar 77,9851%.

DAFTAR PUSTAKA

- [1] S. Sutomo and N. H. Purwanto, "Pengaruh Konsumsi Tisane Daun Belimbing Wuluh terhadap Perubahan Kadar Gula dalam Darah pada Penderita Diabetes Mellitus Tipe 2," *J. Keperawatan*, vol. 16, no. 1, pp. 1–15, 2023.
- [2] P. Sagita, E. Apriliana, S. Mussabiq, and

- T. Soleha, "Pengaruh Pemberian Daun Sirsak Terhadap Penyakit Diabetes," *J. Med. Utama*, vol. 3, no. 1, pp. 1266–1272, 2020.
- [3] N. Latifah, D. Herdiansyah, and A. A. Nasyithoh, "Edukasi Kesehatan Diabetes Mellitus Di Rw.004 Kelurahan Benda Baru Kota Tangerang Selatan," *AS-SYIFA J. Pengabd. dan Pemberdaya. Kesehat. Masy.*, vol. 1, no. 1, p. 23, 2020, doi: 10.24853/assyifa.1.1.23-27.
- [4] S. Gemini and R. Natalia, "Hubungan Tekanan Darah dan Obesitas Sentral dengan Kadar Gula Darah pada Lansia Penderita Diabetes Melitus Tipe II," *J. Keperawatan Muhammadiyah*, vol. 8, no. 4, 2023.
- [5] R. J. Tiurma and Syahrizal, "Obesitas Sentral dengan Kejadian Hiperglikemia pada Pegawai Satuan Kerja Perangkat Daerah," *Higeia J. Public Heal. Res. Dev.*, vol. 5, no. 3, pp. 227–238, 2021.
- [6] D. Hardianto, "Telaah Komprehensif Diabetes Melitus: Klasifikasi, Gejala, Diagnosis, Pencegahan, Dan Pengobatan," *J. Bioteknol. Biosains Indones.*, vol. 7, no. 2, pp. 304–317, 2021, doi: 10.29122/jbbi.v7i2.4209.
- [7] D. A. Hestiana, "Faktor-Faktor yang Berhubungan dengan Kepatuhan dalam Pengelolaan Diet pada Pasien Rawat Jalan Diabetes Mellitus tipe 2 di Kota Semarang," *J. Heal. Educ.*, vol. 2, no. 2, pp. 138–145, 2017.
- [8] M. S. Fitria *et al.*, "Edukasi Pencegahan Penyakit Diabetes Melitus dan Pemeriksaan Kadar Gula Darah Sewaktu di Panti Asuhan," *J. Inov. Dan Pengabd. Masy. Indones.*, vol. 2, no. 3, pp. 45–48, 2023, doi: 10.26714/jipmi.v2i3.130.
- [9] Balyan, S. Andala, and Y. Akbar, "Hubungan Aktivitas Fisik Dan Kualitas Hidup Pasien Diabetes Melitus," *J. Assyifa' Ilmu Kesehat.*, vol. 8, no. 2, pp. 1–9, 2023, [Online]. Available: file:///C:/Users/asus/Downloads/66._JURNAL+BALYAN_ok[1].pdf.
- [10] S. Anggriani, Y. Mayawati, N. Jalil, and B. Baharuddin, "Pengaruh Latihan Senam Kaki terhadap Kadar Glukosa Darah pada Pasien DM Tipe 2 di Wilayah Kerja Puskesmas Bangkala Kota Makassar," *J. Media Keperawatan Politek. Kesehat. Makassar*, vol. 12, no. 1, pp. 87–92, 2021.
- [11] Priyanto Priyanto, Nengsih Yulianingsih, and Hasim Asyari, "Hubungan Pengetahuan Tentang Diabetes Mellitus Dengan Kepatuhan Menjalani Pengobatan Pada Pasien Diabetes Mellitus Di Kecamatan Kertasemaya Tahun 2021," *J. Pengabd. Ilmu Kesehat.*, vol. 2, no. 1, pp. 17–24, 2022, doi: 10.55606/jpikes.v2i1.337.
- [12] R. Marzel, "Terapi pada DM Tipe 1," *J. Penelit. Perawat Prof.*, vol. 3, no. 1, pp. 51–62, 2021, doi: 10.37287/jppp.v3i1.297.
- [13] A. N. Faida and Y. D. P. Santik, "Kejadian Diabetes Melitus Tipe I pada Usia 10-30 Tahun," *Higeia J. Public Heal. Res. Dev.*, vol. 4, no. 1, pp. 33–42, 2020.
- [14] N. Sundari, F. S. T. Dewi, and M. R. Ikhsan, "Kebiasaan gaya hidup dan obesitas sebagai faktor risiko kejadian diabetes melitus tipe 2 di rumah sakit Kutai Kartanegara," *J. Ber. Kedokt. Masy.*, vol. 32, no. 12, pp. 461–466, 2016, [Online]. Available: <https://doi.org/10.22146/bkm.8337>.
- [15] M. K. Murtiningsih, K. Pandelaki, and B. P. Sedli, "Gaya Hidup sebagai Faktor Risiko Diabetes Melitus," *e-Clinic*, vol. 9, no. 2, pp. 328–333, 2021, doi: <https://doi.org/10.35790/ecl.9.2.2021.32852>.
- [16] S. Gunawan and R. Rahmawati, "Hubungan Usia, Jenis Kelamin dan Hipertensi dengan Kejadian Diabetes Mellitus Tipe 2 di Puskesmas Tugu Kecamatan Cimanggis Kota Depok Tahun 2019," *ARKESMAS (Arsip Kesehat. Masyarakat)*, vol. 6, no. 1, pp. 15–22, 2021, doi: 10.22236/arkesmas.v6i1.5829.
- [17] F. Sijabat, R. Siregar, and T. Sitanggang, "Pemberian Jus Buah Naga pada Lansia yang Menderita DM Tipe 2 di Puskesmas Darussalam," *J. Abdimas Mutiara*, vol. 3, no. 1, pp. 295–306, 2022.
- [18] N. Rita, "Hubungan Jenis Kelamin , Olah Raga Dan Obesitas dengan

- Kejadian Diabetes Mellitus pada Lansia,” *J. Ilmu Kesehat.*, vol. 2, no. 1, pp. 93–100, 2018.
- [19] T. O. Hardiyanti, A. Wurjanto, N. Kusariana, and R. Hestningsih, “Hubungan Jenis Kelamin dan Bidang Studi dengan Praktik Pencegahan Diabetes Mellitus Tipe 2 pada Mahasiswa (Studi pada Mahasiswa Universitas Diponegoro Semarang),” *J. Kesehat. Masy.*, vol. 9, no. 2, pp. 175–179, 2021, doi: 10.14710/jkm.v9i2.28662.
- [20] T. Hidayat and S. Komariah, “Pemilihan Siswa-Siswi Berprestasi Menggunakan Metode Weighted Product (WP) Studi Kasus SMP-Al Fitroh Tangerang,” *J. Tek. Inform. Unis*, vol. 7, no. 2, pp. 159–163, 2019, doi: 10.33592/jutis.v7i2.398.
- [21] A. Pebdika, R. Herdiana, and D. Solihudin, “Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.
- [22] J. Lemantara, “Penerapan Algoritma Naïve Bayes dan ID3 untuk Memprediksi Segmentasi Pelanggan pada Penjualan Mobil,” *J. Technol. Informatics*, vol. 4, no. 1, pp. 31–40, 2022, doi: 10.37802/joti.v4i1.265.
- [23] J. Lemantara, “Rancang Bangun Aplikasi Hipertensi.edu sebagai Media Edukasi dan Diagnosis Penyakit Hipertensi Menggunakan metode Naïve Bayes dengan Laplace Correction,” *INFOTECH J. Inform. Teknol.*, vol. 5, no. 1, pp. 146–160, 2024, doi: 10.37373/infotech.v5i1.1197.
- [24] E. Nasri and A. S. AW, “Aplikasi Seleksi Penentuan Nasabah Untuk Penjualan Barang Secara Kredit Dengan Algoritma K-Nearest Neighbor,” *J. Ilm. Sains dan Teknol. Univ. Baten Jaya*, vol. 4, no. 1, pp. 1–11, 2020.
- [25] I. Luqman Hakim, A. Saputra, M. F. Sahetapi, Y. Ratte, and A. A. Ahoren, “Penerapan Metode K-Nearest Neighbors (KNN) dalam Klasifikasi Golongan SIM di Daerah Kabupaten Manokwari,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3693–3698, 2023, doi: 10.36040/jati.v7i6.8266.
- [26] Y. Yuliska and K. U. Syaliman, “Peningkatan Akurasi K-Nearest Neighbor Pada Data Index Standar Pencemaran Udara Kota Pekanbaru,” *IT J. Res. Dev.*, vol. 5, no. 1, pp. 11–18, 2020, doi: 10.25299/itjrd.2020.vol5(1).4680.
- [27] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 427–434, 2018, doi: 10.25126/jtiik.201854773.
- [28] M. Brilliant, I. A. Nurhasanah, and D. Rahmadaniah, “Perbandingan Algoritma Naive Bayes dan K-Nearest Neighbor untuk Kasifikasi Waktu Tunggu Alumni dalam Memperoleh Pekerjaan (Study KAsus SKMS PGRI 2 Pringsewu),” *SEAT J. Softw. Eng. Technol.*, vol. 3, no. 2, pp. 15–20, 2023, [Online]. Available: <https://journal.institdla.ac.id/index.php/seat/article/view/101%0Ahttps://journal.institdla.ac.id/index.php/seat/article/viewFile/101/91>.
- [29] M. R. Yuliansyah, M. B, and A. Franz, “Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda,” *Adopsi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 08–20, 2022, doi: 10.30872/atasi.v1i1.25.
- [30] N. M. Putry and B. N. Sari, “Komparasi Algoritma KNN Dan Naïve Bayes untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus,” *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, pp. 45–57, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [31] M. Danny and S. Umam, “Penerapan Data Mining Menggunakan Algoritma Fp-Growth Untuk Menganalisa Pola Penjualan Obat (Studi Kasus: Klinik Annisa),” *Pros. SAINTEK Univ. Pelita Bangsa*, vol. 1, no. 1, pp. 159–164, 2022.
- [32] E. Martantoh and N. Yanih, “Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik

- Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan Php Mysql," *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 166–175, 2022, doi: 10.35957/jtsi.v3i2.2896.
- [33] A. B. Mustriyanto, M. Habibi, D. Subekti, and F. Syahrudin, "Perbandingan Metode Decision Tree dan Naive Bayes Classifier pada Analisis Sentimen Pengguna Layanan PT Perusahaan Listrik Negara (PLN)," *Teknomatika J. Inform. dan Komput.*, vol. 15, no. 2, pp. 53–61, 2022, doi: 10.30989/teknomatika.v15i2.1131.
- [34] M. Muchtar, Y. P. Pasrun, R. Rasyid, N. Miftachurohmah, and M. Mardawati, "Penerapan Metode Naïve Bayes dalam Klasifikasi Kesegaran Ikan Berdasarkan Warna pada Citra Area Mata," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, pp. 611–617, 2024, doi: 10.23960/jitet.v12i1.3879.
- [35] R. A. Nolly, A. Fitria, and K. Saputra, "Penerapan Algoritma K-Nearest Neighbors untuk Klasifikasi Fragmen Metagenom Berdasarkan Ekstraksi Fitur K-Mers," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 17, no. 1, pp. 52–56, 2022, doi: 10.30872/jim.v17i1.5779.
- [36] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.