Versi Online tersedia di : https://jurnal.buddhidharma.ac.id/index.php/algor/index

JURNAL ALGOR

|2715-0577 (Online)| 2715-0569 (Print)

Artikel

Penerapan Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naive Bayes

Hosea Adrianus[1], *Desiyanna Lasut*[2]

[1,2] *Univeristas Buddhi Dharma, Teknik Informatika, Banten, Indonesia*

SUBMISSION TRACK

Recieved: Jan 22, 2023
Final Revision: March 12,2023
Available Online: March 24, 2023

KEYWORD

Data Mining, Diabetes, Gejala Awal Diabetes, Prediksi, *Naive Bayes*

KORESPONDENSI

E-mail: hosea.adrianus@gmail.com

A B S T R A K

Diabetes adalah salah satu penyakit berbahaya di dunia dan salah satu penyakit yang menyebabkan banyak orang meninggal dunia. Berdasarkan laporan dari *World Health Organization* (WHO) pada tahun 2019 penyakit diabetes berada di urutan ke-9 dalam penyebab kematian di seluruh dunia dengan total perkiraan mencapai 1,5 juta kematian. Menurut laporan dari *International Diabetes Federation* (IDF) pada tahun 2021 pengidap penyakit diabetes di seluruh dunia mencapai 537 juta orang atau 1 dari 10 orang di dunia mengidap penyakit diabetes. Dengan bahayanya penyakit diabetes, dibutuhkan suatu sistem untuk memprediksi penyakit diabetes berdasarkan gejala-gejala awal yang dirasakan menggunakan bantuan teknologi dan informasi untuk segera ditangani lebih lanjut. Teknik yang dapat digunakan untuk mengatasi permasalahan tersebut adalah Data Mining menggunakan algoritma *Naive Bayes*. Algoritma *Naive Bayes* adalah salah satu algoritma data mining yang ditemukan oleh Thomas Bayes yang menggunakan teknik probabilitas dan statistika untuk memperkirakan probabilitas di masa yang akan datang berdasarkan pengamatan pada masa lalu.. Untuk melakukan penelitian ini, dataset yang digunakan adalah *Early Stage Diabetes Risk Prediction Dataset* yang diambil dari website *UCI Machine Learning*. Hasil dari penelitian menunjukan bahwa penerapan algoritma *Naive Bayes* untuk memprediksi penyakit diabetes berdasarkan gejala-gejala awal menghasilkan akurasi sebesar 87,69% dengan pengujian menggunakan *RapidMiner*. Berdasarkan hasil tersebut peneliti mengimplementasikan algoritma *Naive Bayes* pada sebuah aplikasi berbasis web untuk memprediksi penyakit diabetes berdasarkan gejala-gejala awal.

PENDAHULUAN

Diabetes adalah salah satu penyakit berbahaya di dunia dan salah satu penyakit yang menyebabkan banyak orang meninggal dunia. Berdasarkan laporan dari *World Health Organization* (WHO) pada tahun 2019 penyakit diabetes berada di urutan ke-9 dalam penyebab kematian di seluruh dunia dengan total perkiraan mencapai 1,5 juta kematian[1]. Menurut laporan dari *International Diabetes Federation* (IDF) pada tahun 2021 pengidap penyakit diabetes di seluruh dunia mencapai 537 juta orang atau 1 dari 10 orang di dunia mengidap penyakit diabetes[2]. Di indonesia sendiri menurut laporan dari *International Diabetes Federation* (IDF) pada tahun 2021 pengidap penyakit diabetes diperkirakan mencapai 19,47 juta orang dan menempati posisi kelima dalam kasus pengidap diabetes terbanyak di dunia[2].

Diabetes adalah penyakit kronis yang menyebabkan pengidapnya tidak dapat mengolah dan menyerap kadar gula darah pada sel untuk digunakan sebagai energi. Yang akhirnya kondisi ini menyebabkan penumpukan gula yang berlebih dalam tubuh. Diabetes adalah suatu penyakit ketika tubuh seseorang tidak dapat memproduksi hormon insulin, atau ketika tubuh tidak mampu untuk menggunakan hormon insulin dengan baik. Insulin merupakan hormon yang dihasilkan organ pankreas yang berfungsi untuk menyerap dan mengolah gula darah untuk dijadikan sumber energi. Tanpa adanya insulin, gula darah tidak bisa diolah dan diserap oleh tubuh menjadi sumber energi.

Ada 2 tipe diabetes, yaitu diabetes tipe 1 dan diabetes tipe 2. Diabetes tipe 1 adalah penyakit diabetes dimana tubuh penderitanya tidak dapat menghasilkan hormon insulin. Sehingga terjadinya penumpukan kadar gula darah dalam tubuh yang bisa menyebabkan kerusakan pada organ-organ tubuh. Untuk penyebab diabetes 1 saat ini belum diketahui penyebabnya secara pasti, namun diduga genetik penderita merupakan faktor yang sangat berpengaruh. Sedangkan diabetes tipe 2 adalah tipe diabetes dimana tubuh penderita menjadi mengalami penurunan sensitifitas terhadap hormon insulin, yang menyebabkan insulin tidak dapat diserap dan digunakan dengan baik atau(resistensi insulin). Diabetes tipe 2 merupakan tipe diabetes yang lebih umum diderita oleh masyarakat. Diabetes tipe 2 biasanya terjadi dikarenakan gaya hidup dan bertambahnya usia.

Dengan bahayanya penyakit diabetes, dibutuhkan suatu sistem untuk memprediksi penyakit diabetes berdasarkan gejala-gejala awal yang dirasakan menggunakan bantuan teknologi dan informasi untuk segera ditangani lebih lanjut. Perkembangan teknologi dan informasi saat ini sangat pesat dan sudah memasuki berbagai bidang kehidupan manusia, salah satunya adalah bidang medis. Salah satu teknik yang dapat digunakan dalam teknologi dan informasi di bidang medis adalah Data Mining, yang dapat digunakan untuk mengklasifikasi dan memprediksi suatu penyakit.

Data mining adalah sebuah proses ekstraksi atau penggalian dari suatu data yang besar menggunakan teknik analisis dengan tujuan mendapatkan informasi yang berguna[3]. Salah satu algoritma yang digunakan untuk mengklasifikasi dan prediksi dalam data mining adalah *Naive Bayes*. *Naive Bayes* adalah salah satu algoritma yang ada dalam data mining yang ditemukan oleh Thomas Bayes yang menggunakan teknik probabilitas dan statistika untuk memperkirakan probabilitas di masa yang akan datang berdasarkan pengamatan pada masa lalu[4]. *Naive Bayes* dapat digunakan untuk memprediksi variabel target dengan basis probabilitas. Dalam kasus prediksi penyakit diabetes, variabel target adalah apakah seseorang menderita diabetes atau tidak. Oleh karena itu, *Naive Bayes* dapat digunakan untuk memperkirakan probabilitas seseorang menderita diabetes berdasarkan nilai-nilai dari variabel-variabel yang terkait, seperti usia, indeks massa tubuh, tekanan darah, dan sebagainya. Algoritma *Naive Bayes* digunakan pada penelitian ini dikarenakan peneliti ingin mencari tahu akurasi yang akan dihasilkan oleh algoritma *Naive Bayes* dalam memprediksi penyakit diabetes

I. METODE

Data

Data adalah suatu fakta yang mendeskripsikan suatu peristiwa dan kejadian nyata, yang memerlukan pemrosesan untuk dapat menghasilkan informasi[5]. Data merupakan suatu nilai yang merepresentasikan deksripsi dari suatu objek atau kejadian yang dapat berupa angka, teks, maupun lambang[6].

Dapat disimpulkan bahwa data merupakan sekumpulan nilai-nilai yang bisa berupa angka, teks maupun lambang-lambang yang berasal dari kejadian nyata yang perlu diolah lebih lanjut untuk menghasilkan suatu informasi yang berguna bagi masyarakat.

Data Mining

Data mining adalah proses penggalian informasi dari suatu data yang besar menggunakan teknik statistika, matematika, dan kecerdasan buatan[7].

Data mining adalah suatu kegiatan yang berkaitan dengan pencarian pengetahuan, informasi, atau pola pada data yang jumlahnya banyak atau besar. Pengetahuan yang diperoleh dari data mining dapat digunakan sebagai dasar untuk pengambilan keputusan di masa depan[8].

[9]Data mining dibagi menjadi beberapa jenis, yaitu :

a. Klastering (*Clustering*)
   Klastering data mining adalah suatu proses untuk mengelompokan suatu kelas kedalam beberapa segmen berdasarkan kemiripan atribut-atribut yang dimiliki.
b. Klasifikasi
   Klasifikasi adalah suatu proses menemukan definisi kesamaan karakteristik dalam suatu kelas (*class*). Klasifikasi bertujuan untuk memperkirakan kelas (*class*) dari suatu objek yang label nya belum diketahui.
c. Asosiasi
   Asosiasi adalah proses untuk menganalisa kombinasi item yang muncul bersamaan. Asosiasi biasa disebut analisis keranjang belanja (*Market Basket Analysis*). Asosiasi dapat digunakan untuk mengidentifikasi produk yang sering dibeli bersamaan oleh pelanggan.
d. Estimasi
   Estimasi merupakan suatu proses untuk memperkirakan sebuah nilai yang belum diketahui. Estimasi mirip seperti prediksi namun estimasi lebih mengarah ke arah numerik daripada kategori.
e. Prediksi
   Prediksi adalah suatu proses untuk memperkirakan hasil yang akan datang di masa depan. Prediksi juga mirip dengan estimasi dan klasifikasi. Untuk memprediksi juga dapat menggunakan teknik klasifikasi atau estimasi.

Naive Bayes

Naive Bayes adalah salah satu algoritma yang ada dalam data mining yang ditemukan oleh Thomas Bayes yang menggunakan teknik probabilitas dan statistika untuk memperkirakan probabilitas di masa yang akan datang berdasarkan pengamatan pada masa lalu[4].

Algoritma *Naive Bayes* adalah suatu teknik yang dapat digunakan untuk memprediksi, yang menggunakan basis probabilitas sederhana berdasarkan pada teorema bayes dengan asumsi bahwa tiap fitur tidak bergantung pada fitur yang lainnya atau independensi fitur[10].

Naive Bayes memiliki keunggulan dalam implementasi nya yang mudah dan memberikan hasil yang memuaskan pada banyak kasus. Namun *Naive Bayes* memiliki kekurangan, yaitu mengasumsikan fitur-fitur yang ada tidak terkait satu sama lainnya meskipun dalam kenyataan itu ada.

Berikut adalah perhitungan probabilitasnya :

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

Keterangan :

1) X = Merupakan data yang akan dicari kelas nya

2) P (X) = Merupakan probabilitas X
3) H = Merupakan hipotesis bahwa data X termasuk ke dalam kelas tertentu
4) P (H) = Merupakan probabilitas Hipotesis H (prior probability)
5) P (X|H) = Merupakan probabilitas X berdasarkan kondisi pada hipotesis H
6) P (H|X) = Merupakan probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)

Algoritma

Algoritma adalah sebuah langkah-langkah yang melibatkan serangkaian operasi yang disusun dan diatur secara logis dan sistematis untuk memecahkan masalah dan menghasilkan output tertentu[11]. Algoritma adalah suatu rangkaian dari aksi-aksi yang berurutan yang didefinisikan secara jelas untuk dapat memecahkan masalah dalam rentang waktu tertentu[12].

Prediksi

Prediksi adalah suatu kegiatan yang meramalkan suatu variabel di masa mendatang yang berdasarkan pada data ataupun variabel pada masa lampau[13].
Prediksi adalah sebuah perkiraan yang secara spesifik memprediksi bentuk observasi yang akan terjadi di masa depan. Dalam melakukan prediksi, sangat penting untuk didasarkan pada hasil pengamatan dan pengukuran yang akurat, bukan hanya berdasarkan spekulasi atau tebakan semata. Artinya, prediksi harus didasarkan pada data yang terukur dan teliti, dan tidak hanya berdasarkan tebak-tebakan semata[14].

Diabetes Mellitus

Diabetes adalah penyakit kronis yang menyebabkan pengidapnya tidak dapat mengolah dan menyerap kadar gula darah pada sel untuk digunakan sebagai energi[15]. Yang akhirnya kondisi ini menyebabkan penumpukan gula yang berlebih dalam tubuh. Diabetes adalah suatu penyakit ketika tubuh seseorang tidak dapat memproduksi hormon insulin, atau ketika tubuh tidak mampu untuk menggunakan hormon insulin dengan baik. Insulin merupakan hormon yang dihasilkan organ pankreas yang berfungsi untuk menyerap dan mengolah gula darah untuk dijadikan sumber energi. Tanpa adanya insulin, gula darah tidak bisa diolah dan diserap oleh tubuh menjadi sumber energi.
Gejala diabetes biasanya bervariasi dari setiap pengidapnya, berikut adalah gejala –gejala umum yang biasanya muncul pada pengidap diabetes :

a. Mengalami haus yang berlebihan.
b. Buang air kecil yang berlebihan, khususnya pada malam hari.
c. Merasa kelaparan yang berlebihan.
d. Berat badan turun drastis secara tiba-tiba tanpa sebab yang jelas.
e. Adanya gangguan penglihatan, seperti pandangan kabur.
f. Mudah lelah.
g. Penurunan massa otot.
h. Terdapat keton pada urin. Keton merupakan hasil limbah ketika tubuh tidak dapat mengubah gula menjadi energi yang berasal dari pemecahan otot dan lemak.
i. Sering terkena infeksi pada gusi, kulit, alat kelamin (khususnya pada wanita), serta saluran kemih.

II. PERANCANGAN

Dataset

Untuk melakukan penelitian ini memerlukan sebuah dataset yang akan digunakan sebagai data latih untuk memprediksi penyakit diabetes. Pada penelitian ini menggunakan data sekunder yang diambil dari website penyedia data yaitu UCI Machine Learning Repository yang bernama *Early Stage Diabetes Risk Prediction Dataset*.
Dataset tersebut memiliki 520 record, 17 atribut dan 1 dari atribut tersebut adalah atribut hasil atau kelas. 17 atribut yang ada pada dataset tersebut adalah *Age, Gender, Polyuria, Polydipsia, Sudden Weight Loss, Weakness, Polyphagia, Genital Thrush, Visual Blurring, Itching, Irritability, Delayed*

Healing, *Partial Paresis*, *Muscle Stiffness*, *Alopecia*, *Obesity*, dan *Class*.
Berikut adalah penjelasan dari 17 atribut tersebut :

a. *Age* adalah usia pasien yang menjalani tes. Nilai = 16 – 90.
b. *Gender* adalah jenis kelamin pasien yang menjalani tes. Nilai = *Male* dan *Female.*
c. *Polyuria* adalah apakah pasien mengalami buang air kecil yang berlebihan. Nilai = *Yes* dan *No.*
d. *Polydipsia* adalah apakah pasien mengalami rasa haus yang berlebihan. Nilai = *Yes* dan *No.*
e. *Sudden Weight Loss* adalah apakah pasien mengalami penurunan berat badan secara drastis. Nilai = *Yes* dan *No.*
f. *Weakness* adalah apakah pasien sering merasa kelelahan. Nilai = *Yes* dan *No.*
g. *Polyphagia* adalah apakah pasien mengalami kelaparan yang berlebihan. Nilai = *Yes* dan *No.*
h. *Genital Thrush* adalah apakah pasien mengalami infeksi jamur di bagian genital. Nilai = *Yes* dan *No*.
i. *Visual Blurring* adalah apakah pasien mengalami penglihatan yang kabur. Nilai = *Yes* dan *No*.
j. *Visual Blurring* adalah apakah pasien mengalami penglihatan yang kabur. Nilai = *Yes* dan *No*.
k. *Irritability* adalah apakah pasien sering merasa mudah tersinggung atau marah. Nilai = *Yes* dan *No*.
l. *Delayed Healing* adalah apakah pasien mengalami penyembuhan luka yang lama. Nilai = *Yes* dan *No*.
m. *Partial Paresis* adalah apakah pasien sulit menggerakan beberapa bagian anggota badan. Nilai = *Yes* dan *No*.
n. *Muscle Stiffness* adalah apakah pasien mengalami otot kaku atau kram. Nilai = *Yes* dan *No*.
o. *Alopecia* adalah apakah pasien mengalami kerontokan rambut. Nilai = *Yes* dan *No*.
p. *Obesity* adalah apakah pasien dikategorikan sebagai obesitas (angka BMI diatas 25). Nilai = *Yes* dan *No*.
q. *Class* adalah atribut kelas yang menyatakan apakah pasien mengidap penyakit diabetes atau tidak. Nilai = *Positive* dan *Negative*.

Tabel 1. Tabel Sampel Dataset

Age	Gender	Polydipsia	Sudden weight loss	weakness	Polyphagia	Genital Thrush	Visual Blurring	Itching	Irritability	Delayed Healing	Partial Paresis	Muscle Stiffness	Alopecia	Obesity	Class
40	Male	No	No	Yes	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Positive
58	Male	No	No	Yes	No	No	Yes	No	No	No	Yes	No	Yes	No	Positive
41	Male	Yes	No	Yes	Yes	No	No	Yes	No	Yes	No	Yes	Yes	No	Positive
45	Male	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No	No	Positive
60	Male	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Positive
….	…..	….	….	….	….	….	….	….	….	….	….	….	….	….	….

Transformasi Dataset

1. *Age* :
    a. Dibawah 35 = <=35
    b. 36 sampai 45 = 36 – 45
    c. 46 sampai 55 = 46 – 55
    d. 56 sampai 65 = 56 – 65
    e. Diatas 65 = >65
2. *Gender* :
    a. Male = Pria
    b. Female = Wanita
3. *Polyuria, Polydipsia, Sudden Weight Loss, Weakness, Polyphagia, Genital Thrush, Visual Blurring, Itching, Irritability, Delayed Healing, Partial Paresis, Muscle Stiffness, Alopecia, Obesity :*
    a. Yes = Ya
    b. No = Tidak
4. *Class :*
    a. *Positive* = Positif
    b. Negative = Negatif

Tabel 2. Tabel Sampel Dataset Setelah Ditransformasi

Age	Gender	Polydipsia	Sudden weight loss	weakness	Polyphagia	Genital Thrush	Visual Blurring	Itching	Irritability	Delayed Healing	Partial Paresis	Muscle Stiffness	Alopecia	Obesity	Class
40	Pria	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya	Ya	Ya	Positif
58	Pria	Tidak	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Postiif
41	Pria	Ya	Tidak	Ya	Ya	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya	Ya	Tidak	Positif
45	Pria	Tidak	Ya	Ya	Ya	Ya	Tidak	Ya	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Positif
60	Pria	Ya	Ya	Ya	Ya	Tidak	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Positif
….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….

III. PEMBAHASAN

Perhitungan Algoritma *Naive Bayes*

Tahapan awal perhitungan *Naive Bayes* adalah dengan melakukan pengambilan data training dari *Early Stage Diabetes Risk Prediction Dataset*. Dataset tersebut memiliki 520 record, 17 atribut dan 1 dari atribut tersebut adalah atribut hasil atau kelas. Dataset tersebut diolah untuk dapat mempermudah proses perhitungan. Untuk Penjelasan dataset yang lebih rinci sudah dijelaskan di bab sebelumnya.

Dari data tersebut, tentukan prediksi Class data X jika (*Age* = 36 – 45, *Gender* = Pria, *Polyuria* = Ya, *Polydipsia* = Ya, *Sudden Weight Loss* = Tidak, *Weakness* = Tidak, *Polyphagia* = Tidak, *Genital Thrush* = Tidak, *Visual Blurring* = Tidak, *Itching* = Tidak, *Irritability* = Tidak, *Delayed Healing* = Ya, *Partial Paresis* = Tidak, *Muscle Stiffness* = Tidak, *Alopecia* = Tidak, *Obesity* = Ya)

Berikut adalah perhitungan prediksi menggunakan algoritma *Naive Bayes* :

1. Baca data training
2. Menghitung jumlah kelas / label
   $(\text{Class} = \text{"Positif"}) = 320$
   $(\text{Class} = \text{"Negatif"}) = 200$
   $$P(\text{Positif}) = \frac{320}{520} = 0.615384615$$
   $$P(\text{Negatif}) = \frac{200}{520} = 0.384615385$$
3. Hitung jumlah kasus dengan kelas yang sama
   Age
   $$P(\text{Age} = \text{“36} - \text{45”} \mid \text{Class} = \text{“Positif”}) = \frac{77}{320} = 0.240625$$
   $$P(\text{Age} = \text{“36} - \text{45”} \mid \text{Class} = \text{“Negatif”}) = \frac{61}{200} = 0.305$$

Gender

P(Gender = "Pria" | Class = "Positif") = $\frac{147}{320}$ = 0.459375

P(Gender = "Wanita" | Class = "Negatif") = $\frac{181}{200}$ = 0.905

Polyuria

P(Polyuria = "Ya" | Class = "Positif") = $\frac{243}{320}$ = 0.759375

P(Polyuria = "Tidak" | Class = "Negatif") = $\frac{15}{200}$ = 0.075

Polydipsia

P(Polydipsia = "Ya" | Class = "Positif") = $\frac{225}{320}$ = 0.703125

P(Polydipsia = "Tidak" | Class = "Negatif") = $\frac{8}{200}$ = 0.04

Sudden Weight Loss

P(Sudden Weight Loss = "Ya" | Class = "Positif") = $\frac{132}{320}$ = 0.4125

P(Sudden Weight Loss = "Tidak" |Class = "Negatif") = $\frac{171}{200}$ = 0.855

Weakness

P(Weakness = "Ya" | Class = "Positif") = $\frac{102}{320}$ = 0.31875

P(Weakness = "Tidak" | Class = "Negatif") = $\frac{113}{200}$ = 0.565

Polyphagia

P(Polyphagia = "Ya" |Class = "Positif") = $\frac{131}{320}$ = 0.409375

P(Polyphagia = "Tidak" | Class = "Negatif") = $\frac{155}{200}$ = 0.76

Genital Thrush

P(Genital Thrush = "Ya" | Class = "Positif") = $\frac{237}{320}$ = 0.740625

P(Genital Thrush = "Tidak" | Class = "Negatif") = $\frac{167}{200}$ = 0.835

Visual Blurring

P(Visual Blurring = "Ya" | Class = "Positif") = $\frac{145}{320}$ = 0.453125

P(Visual Blurring = "Tidak" |Class = "Negatif") = $\frac{142}{200}$ = 0.71

Itching

P(Itching = "Ya" | Class = "Positif") = $\frac{166}{320}$ = 0.51875

P(Itching = "Tidak" | Class = "Negatif") = $\frac{101}{200}$ = 0.505

Irritability

P(Irritability = "Ya" | Class = "Positif") = $\frac{210}{320}$ = 0.65625

P(Irritability = "Tidak" | Class = "Negatif") = $\frac{184}{200}$ = 0.92

Delayed Healing

P(Delayed Healing = "Ya" | Class = "Positif") = $\frac{153}{320}$ = 0.478125

P(Delayed Healing = "Tidak" |Class = "Negatif") = $\frac{86}{200}$ = 0.43

Partial Paresis

P(Partial Paresis = "Ya" | Class = "Positif") = $\frac{128}{320}$ = 0.4

P(Partial Paresis = "Tidak" | Class = "Negatif") = $\frac{168}{200}$ = 0.84

Muscle Stiffness

P(Muscle Stiffness = "Ya" | Class = "Positif") = $\frac{185}{320}$ = 0.578125

P(Muscle Stiffness = "Tidak" | Class = "Negatif") = $\frac{140}{200}$ = 0.7

Alopecia

P(Alopecia = "Ya" | Class = "Positif") = $\frac{242}{320}$ = 0.75625

P(Alopecia = "Tidak" | Class = "Negatif") = $\frac{99}{200}$ = 0.495

Obesity

P(Obesity = "Ya" | Class = "Positif") = $\frac{61}{320}$ = 0.190625

P(Obesity = "Tidak" | Class = "Negatif") = $\frac{27}{200}$ = 0.135

4. Kalikan semua nilai hasil dengan data X yang dicari kelas nya

P(X|Class = "Positif")

$$= \begin{pmatrix} 0.615384615 * 0.240625 * 0.459375 * \\ 0.759375 * 0.703125 * 0.4125 * \\ 0.31875 * 0.409375 * 0.740625 * \\ 0.453125 * 0.51875 * 0.65625 * \\ 0.478125 * 0.4 * 0.578125 * \\ 0.75625 * 0.190625 \end{pmatrix}$$

= 0.00185

P(X|Class = "Negatif")

$$= \begin{pmatrix} 0.384615385 * 0.305 * 0.905 * \\ 0.075 * 0.04 * 0.855 * \\ 0.565 * 0.76 * 0.835 * \\ 0.71 * 0.505 * 0.92 * \\ 0.43 * 0.84 * 0.7 * \\ 0.495 * 0.135 \end{pmatrix}$$

= 0.000283

5. Dari hasil prediksi "Positif" dan "Negatif" dibandingkan dan dicari nilai yang paling tinggi sehingga didapatkan hasil dari prediksi tersebut yaitu

P(X | Class = "Positif") = **0.00185**

P(X | Class = "Negatif") = 0.000283

Persentase :

$$P(X \mid Class = \text{“Positif”}) = \left(\frac{0.00185}{(0.00185 + 0.000283)}\right) * 100 = \mathbf{86.7411160}\%$$

$$P(X \mid Class = \text{“Negatif”}) = \left(\frac{0.000283}{(0.00185 + 0.000283)}\right) * 100 = 11.2588840\%$$

Dari kedua hasil tersebut diketahui nilai P(X | Class = “Positif”) lebih besar dari P(X | Class = “Negatif”), maka class dari data X tersebut adalah “**Positif**”.

Evaluasi Confusion Matrix

Model *Naive Bayes* di uji akurasinya dengan *Confusion Matrix* menggunakan aplikasi RapidMiner dengan pengujian *Cross Validation*. Confusion Matrix akan membentuk sebuah matrix yang terdiri dari *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), dan *False Negatif* (FN) sehingga membentuk matrix seperti gambar di bawah ini :

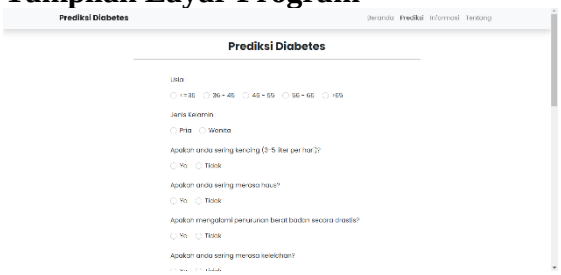
Tabel 3. Confusion Matrix

	TP	TN	Class Precision
Pred. Positif	276	20	93,24%
Pred. Negatif	44	180	80,36%
Class Recall	86,25%	90%	

Berikut adalah cara untuk menghitung akurasi :

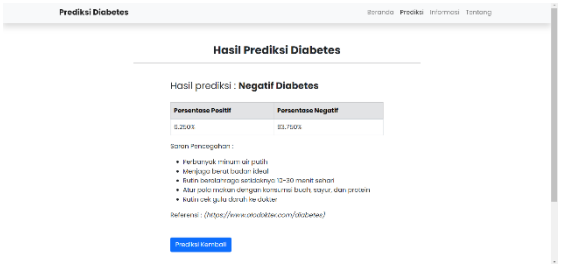
$$Akurasi = \frac{(TP + TN)}{(TP + FP + TN + FN)} = \frac{(276 + 180)}{(276 + 44 + 180 + 20)} = \frac{456}{520} = 0{,}8769 * 100 = 87{,}69\%$$

Tampilan Layar Program



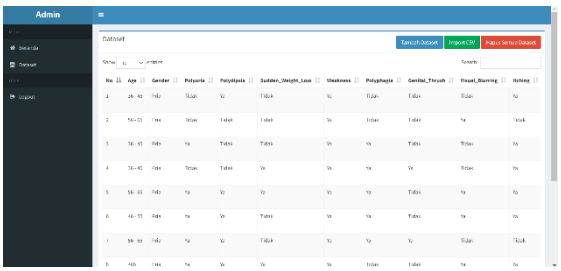
Gambar 2. Tampilan Layar Prediksi Pengguna

Gambar diatas adalah halaman untuk memprediksi diabetes. Kita dapat memilih gejala-gejala yang berhubungan dengan diabetes yang kita rasakan.



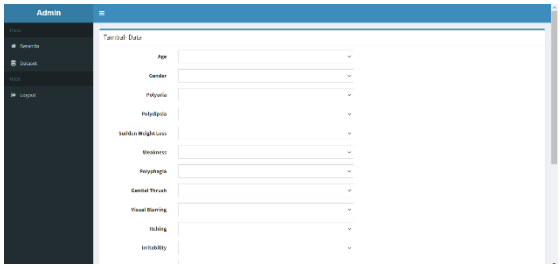
Gambar 3. Tampilan Hasil Prediksi Pengguna

Gambar diatas adalah halaman untuk menampilkan hasil dari prediksi diabetes. Di halaman tersebut menampilkan hasil dari prediksi diabetes, yaitu negatif atau positif. Kemudian ada juga persentase dari negatit dan positif diabetes beserta saran untuk pencegahan ataupun penanganan lebih lanjut



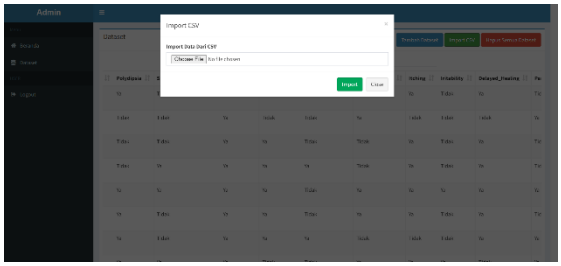
Gambar 4. Tampilan Dataset Admin

Gambar diatas adalah halaman admin untuk melihat dataset yang digunakan dalam aplikasi. Admin dapat memasukkan, mengubah, menghapus dataset dari menu yang telah disediakan.



Gambar 5. Tampilan Tambah dan Ubah Dataset Admin

Gambar diatas adalah halaman admin untuk menambah dan mengubah dataset. Pada halaman ini admin dapat memasukkan atau mengubah dataset berdasarkan atribut-atribut yang ada.



Gambar 6. Tampilan Import Dataset Admin

Gambar diatas adalah halaman admin untuk mengimport dataset. Pada halaman ini admin dapat mengimport dataset yang berjenis file CSV kedalam database aplikasi.

Pengujian Black Box Pengguna

Tabel 3. Pengujian Black Box Pengguna

Scene Pengujian	Skenario	Hasil yang diharapkan	Hasil Uji
Halaman Beranda	Mengklik tombol Prediksi	Mengalihkan halaman ke halaman Prediksi	Sesuai
Halaman Prediksi	Mengisi semua pertanyaan gejala dan mengklik tombol Prediksi	Menampilkan hasil prediksi diabetes berupa class negatif atau positif, persentase, dan saran	Sesuai
Halaman Prediksi	Tidak mengisi semua pertanyaan dan mengklik tombol Prediksi	Mengarahkan ke yang belum diisi	Sesuai
Halaman Informasi	Mengklik Menu Informasi	Menampilkan informasi tentang diabetes seperti definisi diabetes, jenis diabetes, gejala-gejala diabetes, faktor resiko diabetes, pencegahan diabetes	Sesuai
Halaman Tentang	Mengklik Menu Tentang	Menampilkan informasi tentang aplikasi yang dibuat dan algoritma yang digunakan	Sesuai

Pengujian Black Box Admin

Tabel 4. Tabel Pengujian Black Box Admin

Scene Pengujian	Skenario	Hasil yang diharapkan	Hasil Uji
Halaman Login	Memasukkan username dan password sesuai dengan yang ada di database dan mengklik tombol login	Masuk ke halaman Beranda Admin yang menampilkan pesan selamat datang	Sesuai
Halaman Login	Memasukkan username dan password yang tidak ada atau tidak sama dengan yang ada di database dan	Menampilkan pesan *error*	Sesuai

	mengklik tombol login		
Halaman Beranda	Mengklik Menu Dataset	Menampilkan halaman Dataset	Sesuai
Halaman Beranda	Mengklik Menu *Logout*	Melakukan *Logout* dan menampilkan halaman *Login*	Sesuai
Halaman Dataset	Mengklik Menu Beranda	Menampilkan halaman Beranda	Sesuai
Halaman Dataset	Mengklik Menu *Logout*	Melakukan Logout dan menampilkan halaman Login	Sesuai
Halaman Dataset	Klik Tombol Tambah Dataset	Menampilkan form tambah data	Sesuai
Halaman Dataset	Mengklik Tombol Import CSV	Menampilkan form import CSV	Sesuai
Halaman Dataset	Mengklik Tombol Hapus Semua Dataset dan mengklik OK pada pesan konfirmasi	Menghapus semua dataset yang ada di database	Sesuai
Halaman Dataset	Mengklik Tombol Ubah pada data yang dipilih	Menampilkan form ubah data dan atribut-atribut dari data	Sesuai
Halaman Dataset	Mengklik Tombol Hapus pada data yang dipilih dan mengklik tombol OK pada pesan konfirmasi	Menghapus data yang dipilih dari database	Sesuai
Form Import CSV	Memilih dataset yang akan digunakan yang bertipe file CSV dan klik tombol Import	Mengimport data dari file CSV yang dipilih ke dalam database	Sesuai
Form Import CSV	Memilih dataset yang akan digunakan yang bertipe file bukan CSV dan klik tombol Import	Menampilkan pesan *error* import data dan file yang dipilih bukan CSV	Sesuai
Form Tambah Dataset	Mengisi nilai semua atribut-atribut yang ada dan mengklik tombol Simpan	Menambahkan data yang baru diisi ke dalam dataset dan mengalihkan ke Halaman Dataset.	Sesuai
Form Ubah Data	Mengubah nilai atribut-atribut dan mengklik tombol Simpan	Mengubah data dan mengalihkan ke Halaman Dataset	Sesuai

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan oleh penulis, maka dapat ditarik kesimpulan sebagai berikut :

1. Algoritma *Naive Bayes* dapat digunakan untuk memprediksi apakah seseorang terkena penyakit diabetes berdasarkan gejala awal yang dirasakan.
2. Tingkat akurasi yang dihasilkan oleh algoritma *Naive Bayes* dalam memprediksi penyakit diabetes dengan menggunakan dataset "*Early Stage Diabetes Risk Prediction Dataset*" yang didapat dari website *UCI Machine Learning* adalah sebesar 87,69%.

REFERENCES

[1] WHO, "Diabetes - WHO | World Health Organization," 2019. https://www.who.int/news-room/fact-sheets/detail/diabetes#:~:text=In 2019%2C diabetes was the,onset of type 2 diabetes.

[2] IDF, "Diabetes facts & figures," 2021. https://www.idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html#:~:text=The IDF Diabetes Atlas Tenth,and 783 million by 2045.

[3] A. M. Siregar and A. Puspabhuana, *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. 2017.

[4] T. Sanubari, C. Prianto, and N. Riza, *Odol (One Desa One Product Unggulan Online) Penerapan Metode Naive Bayes Pada Pengembangan Aplikasi E-Commerce Menggunakan Codeigniter*. 2020.

[5] T. Sutabri, *Konsep Sistem Informasi*. Yogyakarta: ANDI OFFSET, 2012.

[6] C. A. Pamungkas, *Pengantar dan Implementasi Basis Data*. Yogyakarta: Deepublish, 2017.

[7] M. Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. 2020.

[8] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. 2020.

[9] A. Wanto, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.

[10] A. S. Huda, R. M. Awangga, and Rd. N. S. Fathonah, *Prediksi Penerimaan Pegawai Baru Dengan Metode Naive Bayes*. 2020.

[11] Kani, *Algoritma dan Pemograman*. 2020.

[12] A. Sindar, *STRUKTUR DATA DAN ALGORITMA DENGAN C++*. 2019.

[13] R. Habibi and A. Suryansah, *Aplikasi prediksi jumlah kebutuhan perusahaan*. 2020.

[14] A. Sujana and A. K. Jayadinata, *Pembelajaran Sains di Sekolah Dasar*. UPI Sumedang Press, 2018.

[15] Alodokter, "Diabetes - Gejala, penyebab dan mengobati - Alodokter," 2020.

BIOGRAPHY

Hosea Adrianus, lahir di Tangerang pada tanggal 2 Agustus 2000. Menyelesaikan kuliah strata 1 (S1) Program Studi Teknik Informatika pada tahun 2022 di Universitas Buddhi Dharma.

Desiyanna Lasut, saat ini bekerja sebagai dosen tetap pada Program Studi Teknik Informatika di Universitas Buddhi Dharma.