PREDIKSI DINI RESIKO PENYAKIT DIABETES MENGGUNAKAN JARINGAN SYARAF TIRUAN BACKPROPAGATION

Nita Khoirun Nisa, Dian Ahkam Sani, Muhammad Udin
Teknik Informatika, Universitas Merdeka Pasuruan
Jalan Ir. H. Juanda No. 68 Pasuruan, Indonesia
nitakhoirunnisa08@gmail.com

ABSTRAK

Diabetes mellitus merupakan salah satu penyakit kronis yang ditandai oleh tingginya kadar glukosa dalam darah akibat gangguan pada produksi atau fungsi insulin. Pentingnya deteksi dini terhadap risiko diabetes menjadi krusial untuk mencegah komplikasi serius di masa mendatang. Penelitian ini bertujuan untuk melakukan prediksi awal risiko diabetes menggunakan metode Backpropagation pada model Jaringan Syaraf Tiruan (JST). Data yang digunakan berasal dari sumber terbuka di platform Kaggle, dengan total 100.000 baris data dan 9 fitur, meliputi usia, jenis kelamin, tekanan darah, kadar glukosa, indeks massa tubuh (BMI), kadar HbA1c, hipertensi, penyakit jantung, serta riwayat merokok. Tahapan preprocessing meliputi konversi data kategorikal menggunakan label encoding dan pembagian data menjadi data latih dan data uji. Model JST dibangun dengan dua hidden layer yang masing-masing terdiri dari 10 neuron, dan dilatih menggunakan algoritma backpropagation. Hasil evaluasi menggunakan confusion matrix menunjukkan akurasi sebesar 97%, precision 99%, recall 68%, dan F1-score 81%. Sebagai pembanding, model Naïve Bayes diuji pada data yang sama dan menghasilkan akurasi 90%, precision 46%, recall 64%, serta F1-score 53%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa model JST dengan algoritma Backpropagation memiliki performa yang lebih unggul dalam memprediksi risiko diabetes, khususnya dalam aspek ketepatan klasifikasi data positif.

Kata kunci : *Diabetes, Jaringan Syaraf Tiruan, Naïve bayes, Confusion Matrix*

1. PENDAHULUAN

Diabetes mellitus adalah penyakit kronis yang ditandai dengan tingginya kadar glukosa darah dan gangguan metabolisme lainnya. Menurut penelitian terdahulu menjelaskan bahwa Penyakit diabetes telah menjadi masalah kesehatan yang signifikan di seluruh dunia. Ini adalah penyakit kronis yang mempengaruhi jutaan orang, termasuk 1 perempuan. Penanganan penyakit diabetes menjadi penting karena komplikasi yang terjadi jika tidak di tangani dengan benar oleh karena itu, pengembangan metode yang efektif dalam mendiagnosis penyakit diabetes pada perempuan menjadi penting [1]

Menurut Penelitian sebelumnya juga mengungkapkan bahwa terdapat sejumlah factor yang mempengaruhi risiko terkena diabetes tipe 2 pada lansia di Indonesia, di antaranya adalah jenis pekerjaan, tingkat pendidikan, aktivitas fisik, kebiasaan merokok, kelebihan berat badan, konsumsi alkohol, pola makan yang minim buah dan sayur, serta riwayat hipertensi. [2]

Hasil temuan pada peneliti sebelumnya Diabetes Mellitus merupakan salah satu dari sepuluh penyakit dengan jumlah kasus terbanyak. Pada tahun 2017, tercatat sebanyak 3.623 pasien rawat jalan menderita diabetes, serta 352 pasien rawat inap, yang mayoritasnya adalah perempuan. [3]

Penelitian juga pernah dilakukan dengan mendapatkan hasil yang baik pada pengujian kombinasi jumlah data training dan data testing terbaik trerdapat pada kombinasi jumlah data training sebesar 75% atau 325 data dan jumlah data testing sebesar 25% atau 125 data [4]

Penelitian sebelumnya dapatkan hasil mengenai prediksi tingkat kelulusan mahasiswa perguruan tinggi menggunkan metode backpropagation dapat diambil kesimpulan bahwa mendapatkan hasil akurasi sebesar 92,92% dan menghasilkan nilai MSE 0.177[5]. Ada 3 fase dalam pelatihan BPNN, yaitu fase maju (feed forward), fase mundur (back propagation), dan fase modifikasi bobot [6]

2. TINJAUAN PUSTAKA

2.1. Prediksi

Prediksi mirip dengan klasifikasi dan estimasi, tetapi perbedaannya terletak pada fokusnya pada nilai hasil dimasa depan [7] yang belum diketahui Contohnya, memprediksi kemungkinan seseorang menderita diabetes berdasarkan indikator medis seperti usia, kadar gula, tekanan darah, dan data kesehatan lainnya.

2.2. Diabetes

Diabetes melitus adalah sekumpulan gangguan metabolik yang bervariasi, ditandai oleh peningkatan kadar glukosa darah atau hiperglikemia. Pada kondisi ini, tubuh mengalami kesulitan dalam merespons insulin secara efektif, atau produksi insulin oleh pankreas menurun Jika hiperglikemia berlangsung dalam jangka waktu lama, hal ini dapat memicu komplikasi metabolik akut maupun gangguan saraf kronis [8]. Diabetes melitus merupakan penyakit yang diklasifikasikan menjadi dua jenis, yaitu diabetes tipe 1 dan tipe 2. Diabetes tipe 1 umumnya dipicu oleh faktor genetik, usia, lingkungan, serta faktor internal lainnya, sedangkan diabetes tipe 2

lebih banyak disebabkan oleh gaya hidup tidak sehat dan obesitas. Penanganan diabetes melitus dapat dilakukan melalui beberapa pendekatan, seperti terapi insulin, konsumsi obat antidiabetes, pengobatan alternatif, tindakan pembedahan serta perubahan pola hidup yang sehat, seperti mengonsumsi makanan bergizi dan rutin berolahraga. [9]

2.3. Backpropagation

Jaringan Syaraf Tiruan (JST) merupakan metode yang mampu mengidentifikasi hubungan nonlinier antara variabel beban dan faktor-faktor lain, serta mampu menyesuaikan diri terhadap perubahan data. Salah satu penerapan efektif JST adalah dalam bidang peramalan. Namun, JST dengan satu lapisan tersembunyi memiliki keterbatasan dalam mengenali pola kompleks. Kelemahan ini dapat diatasi dengan menambahkan satu atau lebih hidden layer antara lapisan input dan output. Metode Backpropagation pada Jaringan Syaraf Tiruan (JST) digunakan untuk melatih jaringan agar mampu menyeimbangkan antara kemampuan mengenali pola dari data pelatihan dan kemampuan memberikan respons yang tepat terhadap data baru yang serupa [10].

Pada proses ini pelatihan JST dengan menggunkan metode backpropagation terdiri dari beberapa tahapan yaitu :

a. Inisialisasi bobot secara acak
b. Forward propagation dari input ke output
c. Menghitung error antara output model dan target
d. Backward propagation untuk menghitung delta error dari output ke hidden layer
e. Upadate bobot berdasarkan delta error
f. Ulangi proses hingga error minimum atau tercapai batas iterasi.

Pada penelitian ini menggunakan 2 hidden layer dengan masing – masing terdapat 10 neuron, dengan menggunakan fungsi aktivasi ReLu dan fungsi loss binary crossentropy.

2.4. Naïve Bayes

Naïve Bayes merupakan salah satu metode yang bisa diterapkan dalam proses data mining dengan melakukan pengklasifikasian ke kelas yang dibutuhkan. Naïve Bayes dapat menentukan perbedaan pada data yang terdapat pada sebuah pola atau model [11].

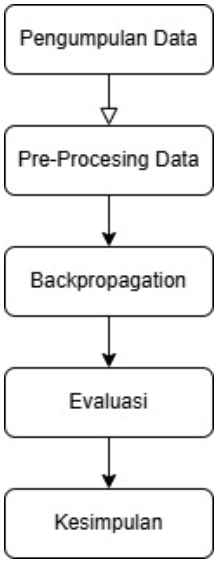
Naive Bayes adalah sebuah algoritma klasifikasi yang berlandaskan pada Teorema Bayes, dengan asumsi dasar bahwa setiap fitur bersifat independen secara kondisional. Dengan kata lain, kemunculan suatu fitur dianggap tidak memiliki keterkaitan dengan kemunculan fitur lainnya dalam proses prediksi.[12]

Pada tahapan naïve bayes ini terdiri dari beberapa tahapan, yaitu :

a. Pisahkan data latih dan data uji
b. Menentukan jumlah data uji
c. Latih menggunakan metode naive bayes
d. Prediksi pada adata uji
e. Lalu menampilkan hasil dari penelitian

3. METODE PENELITIAN

Tahapan dalam penelitian ini dilakukan sesuai dengan alur yang tertera pada gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Pengumpulan Data

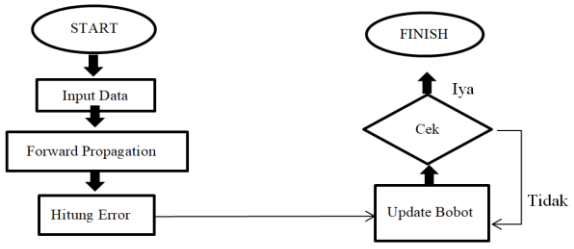
Tahap ini dilakukan untuk mencari refrensi dari penelitian terkait dan mencari data yang akan digunakan untuk penelitian**.**

3.2. Pre-Procesing Data

Tahap kedua ini dilakukan untuk mengetahui apakah data yang didapatkan dapat digunakan untuk penelitian serta memperbaiki apabila ada terjadinya kesalahan pada data**.**

3.3. Penerapan Backpropagation

Pada tahap ketiga ini dilakukan penerapan model jaringan syaraf tiruan (JST) dengan algoritma backpropagation. Pada gambar 2 akan dijelaskan proses yang akan dilakukan dengan menggunakan metode backpropagaton**.**



Gambar 2. Alur backpropagation

Seperti yang dilihat pada gambar 2, tahap pertama yaitu dengan menginisialisasikan parameter dan bobot awal dari jaringan syaraf tiruan yang dimulai secara acak, lalu memasukan parameter bobot awal dari jaringan syaraf tiruan secara acak,

lalu masukan data atau fitur – fitur pada data seperti Jenis kelamin, Umur, Hipertensi, Penyakit jantung, Riwayat merokok, BMI, Hba1c dan Glukosa lalu dilakukan forward propagation untuk menghitung data secara berurutan mulai dari 8 layer, 2 hidden layer dan 1 output yang dijadikan prediksi awal dari model. Lalu hitung selisi dari output dan target selisih ini dihitung sebagai eror yang akan digunakan untuk memperbaiki bobot jaringan, selanjutnya dilakukan Backward propagation yaitu proses yang dilakukan mundur dari output menuju input delta error dihitung dari output layer dan kemudian disebarkan ke hidden layer. Langkah ini melibatkan turunan dari fungsi aktivasi lalu dilakukan update bobot pada jaringan diperbarui berdasarkan nilai delta error yang telah dihitung lalu tahap terakhir dilakukann pengecekan pada error jika masih terdapat eror maka akan terus mengulang hasil yang makin mendekati target

3.4. Evaluasi

Tahap keempat dalam penelitian ini berfokus pada proses evaluasi terhadap hasil prediksi yang telah dihasilkan oleh model. Evaluasi ini bertujuan untuk mengukur tingkat keberhasilan model dalam mengklasifikasikan data secara tepat. Pada tahap ini digunakan metode Confusion Matrix, yaitu alat evaluasi yang menyajikan informasi rinci mengenai performa klasifikasi, termasuk jumlah prediksi yang tepat dan keliru pada masing-masing kelas. Melalui Confusion Matrix, dapat dihitung sejumlah metrik penting seperti akurasi, precision, recall, dan F1-score, yang digunakan sebagai acuan dalam menilai efektivitas model.Setelah itu, hasil evaluasi dari model utama, yakni Jaringan Syaraf Tiruan dengan algoritma Backpropagation, dibandingkan dengan model pembanding, yaitu Naïve Bayes. Tujuan dari perbandingan ini adalah untuk menilai keunggulan relatif dari masing-masing metode yang digunakan. Dengan demikian, peneliti dapat mengidentifikasi metode mana yang memberikan performa terbaik dalam memprediksi kemungkinan seseorang menderita diabetes berdasarkan dataset yang tersedia.

Tabel 1. Komponen Confusion Matrix

Kelas Benar	Diabetes	Tidak Diabetes
Diabetes	True Positif (TP)	False Negatif (FN)
Tidak Diabetes	False Negatif (FN)	True Negatif (TN)

Salah satu metode yang lebih efisien dalam menilai performa klasifikasi adalah dengan menggunakan confusion matrix. Confusion matrix merupakan tabel yang menunjukkan jumlah data uji yang berhasil diklasifikasikan dengan benar maupun yang salah, sehingga membantu dalam menilai tingkat akurasi dari sistem klasifikasi. Melalui confusion matrix, performa sistem klasifikasi dapat dianalisis secara lebih rinci, serta mempermudah dalam mengidentifikasi letak kesalahan klasifikasi yang terjadi.[13]

$$Precision_i = \frac{TP_i}{FP_i+TP_i} \quad (1)$$

Precision digunakan untuk mengetahui seberapa banyak prediksi yang benar di antara semua prediksi

$$Recall_i = \frac{TP_i}{FN_i+TP_i} \quad (2)$$

Fungsi recall yaitu untuk mengetahui sensitivitas seberapa banyak data positif yang berhasil diprediksi dengan benar

$$Accuracy = \frac{\sum_i TP_i}{\sum_l^L \sum_k^L E_{lk}} \quad (3)$$

Akurasi dugunakan untuk mengetahui berpa banyak prediksi yang benar dibandingkan dengan seluruh data yang diuji.

$$F - \ Score = 2\ \frac{Precision \times Recall}{Precission+Recall} \quad (4)$$

F 1 - Score adalah jumlah rata – rata dari precision dan recall

4. HASIL DAN PEMBAHASAN

4.1. Import Dataset

Langkah pertama adalah import library yang dibutuhkan, seperti Pandas. Library Pandas (import pandas as pd) digunakan untuk memanipulasi dan menganlisa data, Label Encording (From sklearn.prepocessing import LabelEncording) adalah fungsi untuk mengubah data kategori ke numerik digunakan untuk mengubah data kategori ke murnerik train_test_split (From sklearn.model_selection import train_test_split) adalah fungsi untuk membagi data latih dan data uji. Sklearn.neural_networknimport MLPlassifier adalah fungsi untuk pemodelan klasifikası menggunkan metode neural network report dan confusion matrix (From sklearn prepocessing import classification report, confusion natrix) adalah fungsi untuk menampilkan hasil klasifikasi dan confusion matrix setelah mengimport libary dan beberapa fungsi yang digunakan untuk pemodelan. Lalu memasukkan data yang akan digunakan.

4.2. Analisa Data

Pada penelitian ini menggunakan data yang bersumber dari kaggle (https://www.kaggle.com/datasets/717451f4573e9c3d8963e626808ade67a89fbd2ee660be008c1aba16b2a2345a) yang berjumlah 100.000 data dengan 8 variabel seperti, Jenis kelamin, umur, hipertensi, riwayat jantung, riwayat merokok, BMI, level Hba1c dan glukosa.

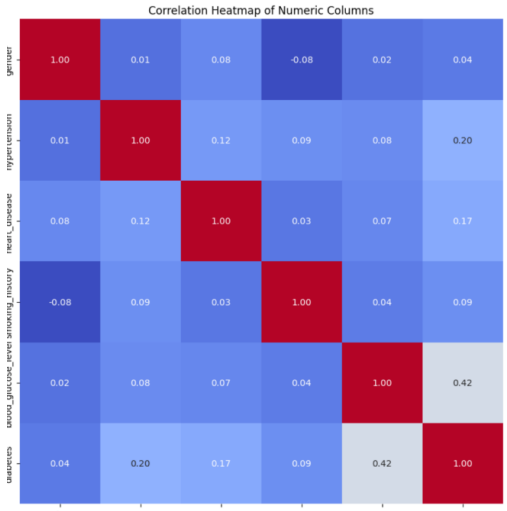
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 9 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   gender               100000 non-null  object
 1   age                  100000 non-null  float64
 2   hypertension         100000 non-null  int64
 3   heart_disease        100000 non-null  int64
 4   smoking_history      100000 non-null  object
 5   bmi                  100000 non-null  float64
 6   HbA1c_level          100000 non-null  float64
 7   blood_glucose_level  100000 non-null  int64
 8   diabetes             100000 non-null  int64
dtypes: float64(3), int64(4), object(2)
memory usage: 6.9+ MB
```

Gambar 3. Informasi Data Set

Lalu dilakukan analisa data yang merupakan proses dimana mengolah dataset dengan melihat informasi dataset seperti tipe data untuk setiap kolom dan distribusi data, memeriksa missing value, melihat korelasi data menggunakan grafik. Berikut adalah langkah proses menganalisa data yang akan digunakan dalam penelitian ini. Pada gambar 4 adalah informasi dataset.

Tahapan awal dalam eksplorasi data dimulai dengan penyajian informasi umum terkait dataset yang digunakan. Dari hasil pemeriksaan, diketahui bahwa dataset ini terdiri atas 100.000 baris data lalu dibagi menjadi 2 yaitu 80% untuk pelatihan dan 20% untuk pengujian tanpa adanya missing value sehingga dapat langsung digunakan dalam proses analisis tanpa memerlukan tahap pembersihan atau imputasi data tambahan. Dataset ini mencakup 9 fitur utama yang berperan sebagai variabel input dalam pemodelan prediksi penyakit diabetes. Sebagai bagian dari proses eksplorasi, dilakukan analisis korelasi antar fitur guna memahami sejauh mana hubungan linier antara masing-masing variabel. Hasil analisis ini divisualisasikan melalui Gambar 4, dalam bentuk heatmap korelasi, yang menunjukkan kekuatan dan arah hubungan antar fitur. Dari visualisasi tersebut, dapat diidentifikasi fitur-fitur yang saling berkorelasi kuat, lemah, maupun tidak berkorelasi sama sekali. Informasi ini sangat berguna dalam proses seleksi fitur, pemahaman struktur data, serta untuk meningkatkan efektivitas dan akurasi model prediktif yang akan dibangun.

Pada tabel korelasi gambar tersebut menunjukan kolom Glukosa memiliki korelasi cukup kuat dengan diabetes (0,42) menunjukan bahwa semakin tinggi kadar glukosa maka semakin besar kemungkinan bahwa seorang menderita diabetes. Hipertensi dan riwayat jantung memiliki korelasi lemah terhadap diabetes dengan masing – masing mendapat nilai (0.20 dan 0.17) Jenis kelamin, riwayat merokok, riwayat jantung memiliki korelasi rendah sekali atau hamper nol terhadap variabel lain, menunjukan bahwa mereka mungkin kurang berpengaruh secara langsung terhadap target atau diabetes.



Gambar 4. Korelasi data antar fitur

4.3. Procesing Data

Tahapan ini merupakan bagian dari proses preprocessing dataset, yang berperan penting dalam mengolah data mentah agar layak digunakan dalam analisis dan pelatihan model machine learning. Dalam penelitian ini, data yang digunakan telah lengkap tanpa adanya nilai yang hilang (*missing value*), sehingga tidak diperlukan langkah tambahan seperti imputasi atau penanganan data kosong. Fokus utama pada tahap ini adalah melakukan konversi data kategorikal menjadi bentuk numerik, karena sebagian besar algoritma pembelajaran mesin hanya dapat bekerja secara optimal dengan data dalam format angka. Proses ini dilakukan melalui teknik label encoding data yang akan dirubah seperti jenis kelamin dan riwayat merokok yang datanya masih menggunakan data nyata dan berbentuk kategorikal seperti contoh data set yang akan di proses menjadi data numerik contoh datanya yaitu pada tabel 2.

Tabel 2. Contoh dataset preprocesing

Jenis kelamin	Umur	Hipertensi	Riwayat Jantung	Riwayat Merokok	BMI	Level Hba1c	Glukosa	Diabetes
Female	80	0	1	Never	25. 19	06.06	140	0
Male	54	0	0	No Info	27 32	06.06	80	0
Male	28	0	0	Never	27.19	05.07	158	0
Female	36	0	0	Current	28.19	05.00	155	0
Male	76	1	1	Current	29.19	04.08	155	0
Female	20	0	0	Never	27.32	06.06	85	0
Female	44	0	0	Never	19.31	06.05	200	1
Female	79	0	0	No Info	23 86	05.07	85	0
Male	42	0	0	Never	33.64	04.08	145	0
Female	32	0	0	Never	26.19	05.00	100	0

Pada dataset seperti pada gambar diatas perlu dilakukan label encoder pada dataset yang bertujuan untuk merubah nilai – nilai yang kategorikal menjadi bentuk nilai numerik. Pada gambar 6 menunjukan bahwa kolom (Jenis kelamin dan riwayat merokok) masih dalam bentuk kategorikal, maka perlu dilakukan label encoding guna untuk mempermudah penelitian. Misalnya “ Female” dirubah menjadi 1 dan “ Male “ dirubah menjadi 0, maka hasil dari Label Encoder seperti pada tabel 3

Tabel 3. Hasil proses label encoder

Jenis kelamin	Umur	Hipertensi	Riwayat Jantung	Riwayat Merokok	BMI	Level Hba1c	Glukosa	Diabetes
1	80	0	1	4	25 19	06.06	140	0
0	54	0	0	0	27 32	06.06	80	0
0	28	0	0	4	27.19	05.07	158	0
1	36	0	0	1	28.19	05.00	155	0
0	76	1	1	1	29.19	04.08	155	0
1	20	0	0	4	27.32	06.06	85	0
1	44	0	0	4	19.31	06.05	200	1
1	79	0	0	0	23 86	05.07	85	0
0	42	0	0	4	33.64	04.08	145	0
1	32	0	0	4	26 19	05.00	100	0

Mengkonversi data yang awalnya bersifat kategorikal ke dalam format numerik menjadikan dataset lebih siap digunakan dalam analisis dan pelatihan model machine learning. Langkah ini sangat penting karena mayoritas algoritma pembelajaran mesin, seperti Jaringan Syaraf Tiruan dan Naïve Bayes, hanya dapat memproses data dalam bentuk angka untuk melakukan perhitungan secara optimal. Proses konversi ini juga berperan dalam mengurangi kemungkinan kesalahan interpretasi oleh algoritma, sekaligus meningkatkan efisiensi dan akurasi dalam pengolahan data. Selain itu, data numerik mempermudah penerapan teknik pra-pemrosesan seperti normalisasi, transformasi fitur, hingga evaluasi model. Oleh karena itu, mengubah data kategorikal menjadi numerik merupakan tahap awal yang esensial dalam membangun penelitian yang lebih sistematis dan tepat sasaran.

Setelah dilakukan preprocessing dengan label encoding terhadap data kategorikal seperti jenis kelamin dan riwayat merokok, seluruh fitur dinormalisasi menggunakan teknik Min-Max Scaling untuk menyesuaikan rentang nilai antar variabel. Dataset kemudian dibagi menggunakan fungsi train_test_split dengan rasio 80% data latih dan 20% data uji. Model Jaringan Syaraf Tiruan (JST) dilatih menggunakan library MLPClassifier dari Scikit-Learn dengan parameter hidden_layer_sizes=(10,10), activation='relu', solver='adam', dan max_iter=1000.

4.4. Permodelan JST Model Backpropagation

Untuk memulai proses perhitungan pada penelitian menggunakan jst backpropagation perlu menentukan arsitektur jaringan. Pada penelitian ini arsitektur jaringan memiliki bentuk 8 lapisan layer yang terdiri dari input layer, 2 hidden layer dan 1 output layer. Jumlah neuron output layer ini didapat berdasarkan target yang ingin

Tabel 4. Arsitektur jaringan

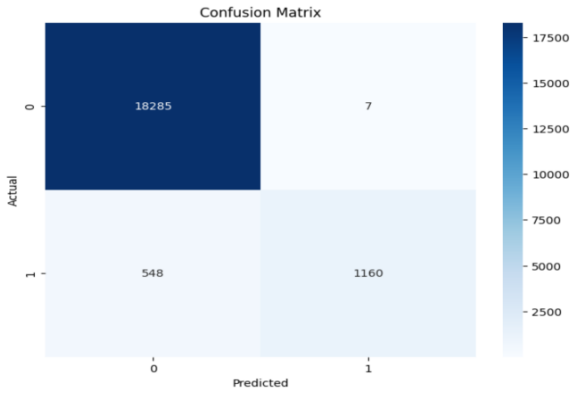
Nama Layer	Jumlah Neuron	Fungsi Aktivitasi
Input	1	-
Hidden Layer 1	10	ReLu
Hidden Layer 2	10	ReLu
Output	1	-

Pada gambar 5 ditunjukkan secara detail arsitektur jaringan dan fungsi aktivasi yang digunakan pada setiap layer pada penelitian ini.

	precision	recall	f1-score	support
0	0.97	1.00	0.99	18292
1	0.99	0.68	0.81	1708
accuracy			0.97	20000
macro avg	0.98	0.84	0.90	20000
weighted avg	0.97	0.97	0.97	20000

Gambar 5. Penerapan model JST

Berdasarkan hasil yang sudah didapatkan terdapat Precision ketepatan prediksi positif (seberapa banyak yang diprediksi benar – benar positif) kelas 0 mendapatkan 97% dari prediksi negative benar dan kelas 1 mendapatkan 99% dari prediksi positif benar. Recall (Kemampuan model menemukan seluruh data positif) pada kelas 0 mendaptkan hasil 100% dari data negative ditemukan semua dan pada kelas 1 hanya mendaptkan 68% dari data positif yang berhasil dikenali.



Gambar 6. Confusion matrix JST

F1-score (Rata – rata harmonic dari precision dan recall pada kelas 0 mendapatkan sangat baik 99% dan pada kelas 1 cukup 0.81% dan masih bias ditingkatkan lalu untuk Support atau jumlah data data sebenarnya di tiap kelas data sangat tidak seimbang terlihat pada kelas 0 jauh lebih banyak dari kelas 1. Accuracy 0.97% model memprediksi dengan benar 97% dari total 20.000 data uji, Rata – rata tanpa mempertimbangkan proporsi setiap kelas. Precision

0.98, recall 0.84 yang menunjukan ketimpangan kinerja antar kelas.Setelah permodelan menggunakan JST dilakukan evaluasi menggunkan confusion matrix. Evaluasi confusion matrix ditampilkan pada gambar 6, hasil evaluasi model menggunakan confusion matrix yang terdiri dari True Negative (TN), True Positive (TP), False Negative (FN), dan False Positive (FP). Masing-masing elemen ini merepresentasikan hasil prediksi model dalam menentukan tingkat akurasi klasifikasi pada data uji. Confusion matrix memberikan gambaran rinci mengenai kesalahan prediksi serta distribusi data aktual pada setiap kelas, sehingga dapat dimanfaatkan untuk menghitung metrik evaluasi seperti precision, recall, dan F1-score. pada tabel terbut menunjukan bahwa True Positif (TP) sejumlah 1.160 orang, False Negative (FN) sejumlah7 orang, False Positive (FP) sejumlah 548 orang, True Negative (TN) sejumlah 18.285 orang .

4.5. Permodelan Naïve Bayes

Dalam penelitian ini, metode Naïve Bayes digunakan sebagai alat pembanding untuk menilai keunggulan algoritma Jaringan Syaraf Tiruan (JST) dengan backpropagation dalam memprediksi penyakit diabetes. Tujuan dari penggunaan metode ini adalah untuk memberikan penilaian yang lebih objektif terhadap performa JST, khususnya dalam hal akurasi, precision, recall, dan F1-score. Agar evaluasi berjalan adil dan konsisten, kedua algoritma diuji menggunakan dataset yang sama, sehingga perbedaan hasil yang muncul benar-benar mencerminkan kapabilitas masing-masing dalam memproses dan menganalisis data. Hasil pengujian performa Naïve Bayes ditampilkan pada Gambar 7, yang mengilustrasikan seberapa efektif metode ini dalam memprediksi potensi diabetes berdasarkan fitur-fitur medis yang tersedia. Melalui perbandingan visual tersebut, dapat diketahui kelebihan maupun kekurangan Naïve Bayes dibandingkan dengan JST Backpropagation..

	precision	recall	f1-score	support
0	0.96	0.93	0.95	18292
1	0.46	0.64	0.53	1708
accuracy			0.90	20000
macro avg	0.71	0.78	0.74	20000
weighted avg	0.92	0.90	0.91	20000

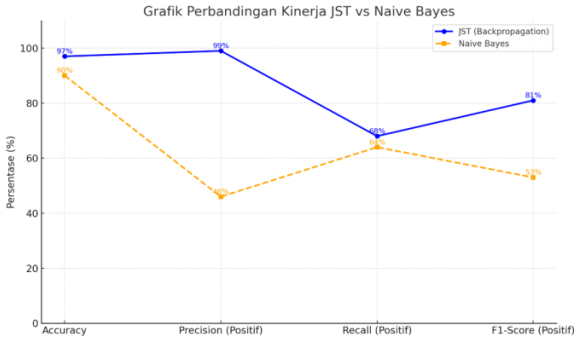
Gambar 7. Penerapan model naïve bayes

Gambar tersebut menunjukkan hasil evaluasi kinerja model klasifikasi. Model memiliki akurasi sebesar 90%, yang artinya dari 20.000 data uji , sebanyak 18.000 data diklasifikasikan dengan benar. Namun, performanya tidak seimbang antar kelas.Untuk kelas 0 (mayoritas), model bekerja sangat baik: Precision: 96%, Recall: 93%, F1-score: 95% Ini berarti model sangat akurat dalam mengenali kelas ini.

Untuk kelas 1 (minoritas), performa jauh lebih rendah: Precision: 46%, Recall: 64%, F1-score: 53% Ini menunjukkan model cukup banyak salah dalam mengenali data kelas 1. Secara keseluruhan, meskipun akurasi tinggi, model kurang baik dalam mengenali kelas 1. Ini bisa disebabkan karena jumlah data kelas 1 jauh lebih sedikit dari kelas 0 (1.708 vs 18.292), sehingga model cenderung mengabaikannya.

4.6. Perbandingan Hasil Pengujian Permodelan

Berdasarkan hasil pengujian yang telah dilakukan menggunakan dua metode, yaitu Jaringan Syaraf Tiruan (JST) dengan algoritma Backpropagation dan algoritma Naïve Bayes, diperoleh hasil evaluasi performa model yang menunjukkan perbedaan signifikan di antara keduanya. Perbandingan hasil ini mencakup metrik-metrik seperti akurasi, precision, recall, dan F1-score. Untuk mempermudah pemahaman dan interpretasi terhadap perbedaan performa tersebut, hasil evaluasi divisualisasikan dalam bentuk grafik garis. Visualisasi tersebut ditampilkan pada Gambar 8, yang memperlihatkan perbandingan kinerja masing-masing metode secara lebih jelas dan informatif. Dengan demikian, pembaca dapat melihat bagaimana performa JST Backpropagation secara umum lebih unggul dibandingkan Naïve Bayes, terutama pada aspek precision dan F1-score



Gambar 8. Hasil perbandingan

Model Jaringan Syaraf Tiruan (JST) berhasil mencapai akurasi sebesar 97%, lebih unggul dibandingkan metode Naive Bayes yang hanya memperoleh 90%. Hal ini menunjukkan bahwa JST secara keseluruhan lebih efektif dalam mengklasifikasikan data, baik untuk kelas negatif maupun positif. Maka JST menghasilkan tingkat kesalahan yang lebih rendah secara umum. Precision menunjukkan seberapa akurat model dalam memprediksi kasus positif (penderita diabetes). JST mencatat precision sebesar 99%, yang berarti hampir seluruh prediksi positifnya benar. Sebaliknya, Naive Bayes hanya mencapai precision 46%, yang mengindikasikan bahwa lebih dari separuh prediksi positifnya keliru (false positive). Menghasilkan JST lebih cocok untuk digunakan dalam proses diagnosis karena kemungkinan memberikan hasil positif palsu sangat kecil. Lalu pada Recall menggambarkan sejauh mana model mampu mendeteksi kasus positif yang sebenarnya. JST memperoleh nilai recall

sebesar 68%, sedikit lebih tinggi dibandingkan dengan Naive Bayes yang mencatat 64%. Meskipun perbedaan tidak terlalu besar, JST menunjukkan kemampuan yang sedikit lebih baik dalam mendeteksi pasien yang benar-benar mengidap diabetes. Keduanya belum optimal dalam mendeteksi seluruh kasus positif, namun JST menunjukkan sensitivitas yang sedikit lebih tinggi. F1-score, sebagai gabungan antara precision dan recall, menunjukkan bahwa JST memperoleh skor 81%, jauh di atas Naive Bayes yang hanya mencapai 53%. Ini menunjukkan bahwa JST mampu menjaga keseimbangan antara ketepatan dan kelengkapan dalam klasifikasi. Menghasilakan JST secara signifikan lebih baik dalam memberikan hasil yang akurat dan menyeluruh

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, model Jaringan Syaraf Tiruan (JST) dengan algoritma Backpropagation menunjukkan performa yang lebih baik dibandingkan metode Naïve Bayes dalam memprediksi risiko diabetes, dengan akurasi Jaringan Syaraf Tiruan sebesar 97%, precision 99%, recall 68%, dan F1-score 81%. Sementara Naïve Bayes hanya mencapai akurasi 90%, precision 46%, recall 64%, dan F1-score 53%. Perbedaan ini menunjukkan bahwa JST lebih akurat dalam mengklasifikasikan kasus positif.

Saran untuk penelitian selanjutnya dapat memperbaiki kemampuan deteksi terhadap kelas minoritas dengan menggunakan metode oversampling atau pengaturan bobot kelas. Di samping itu, penggunaan algoritma lain seperti Random Forest atau SVM juga layak untuk dieksplorasi sebagai pembanding.

DAFTAR PUSTAKA

[1] A. Veronica Agustin and A. Voutama, "Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1002–1007, 2023, doi: 10.36040/jati.v7i2.6808.

[2] F. Milita, S. Handayani, and B. Setiaji, "Kejadian Diabetes Mellitus Tipe II pada Lanjut Usia di Indonesia (Analisis Riskesdas 2018)," *J. Kedokt. dan Kesehat.*, vol. 17, no. 1, p. 9, 2021, doi: 10.24853/jkk.17.1.9-20.

[3] C. A. Rahayu, "PREDIKSI PENDERITA DIABETES MENGGUNAKAN METODE NAIVE BAYES," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, Aug. 2023, doi: 10.23960/jitet.v11i3.3055.

[4] D. A. Sani and M. Z. Sarwani, "Koreksi Jawaban Esai Berdasarkan Persamaan Makna Menggunakan Fasttext dan Algoritma Backpropagation," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 2, pp. 92–111, 2022, doi: 10.23887/janapati.v11i2.49192.

[5] Y. Muamar and A. Muhajirin, "Penerapan Jaringan Saraf Tiruan Dengan Metode Backpropagation Untuk Memprediksi Tingkat Kelulusan Mahasiswa Perguruan Tinggi," *Digit. Transform. Technol.*, vol. 4, no. 1, pp. 214–224, 2024, doi: 10.47709/digitech.v4i1.3810.

[6] S. Auliana, S. Mahrojah, G. Dwiki, and P. Aryono, "Enhanced Facial Expression Recognition Through a Hybrid Deep Learning Approach Combining ResNet50 and ResNet34 Models," vol. 4, no. 6, pp. 2676–2685, 2024, doi: 10.30865/klik.v4i6.1874.

[7] P. A. Duran, A. V. Vitianingsih, M. S. Riza, A. L. Maukar, and S. F. A. Wati, "Data Mining Untuk Prediksi Penjualan Menggunakan Metode Simple Linear Regression," *Teknika*, vol. 13, no. 1, pp. 27–34, 2024, doi: 10.34148/teknika.v13i1.712.

[8] E. E. Mustofa, J. Purwono, and Ludiana, "Penerapan Senam Kaki Terhasap Kadar Glukosa Darah Pada Pasien Diabetes Melitus Di WIlayah Kerja Puskesmas Purwosari Kec. Metro Utara," *J. Cendikia Muda*, vol. 2, no. 1, pp. 78–86, 2022.

[9] J. Biologi *et al.*, "Diabetes Melitus: ReviewEtiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan," 2021. [Online]. Available: http://journal.uin-alauddin.ac.id/index.php/psb

[10] D. Andini, "Jaringan Syaraf Tiruan Untuk Klasifikasi Penyakit Demam Menggunakan Algoritma Backpropagation," *Bull. Artif. Intell.*, vol. 2, no. 1, pp. 86–99, 2023, doi: 10.62866/buai.v2i1.80.

[11] B. R. Tsania Dzulkarnain, Dian Eka Ratnawati, "Penggunaan Metode Naïve Bayes Classifier Pada Analisis the Use of the Naïve Bayes Classifier Method in Sentiment Analysis of the Community ' S Assessment of Hospital Services in," vol. 10, no. 7, pp. 1453–1460, 2023, doi: 10.25126/jtiik.2024117979.

[12] E. Martantoh and N. Yanih, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadiaan 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.

[13] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.