

PENERAPAN BACKPROPAGATION JARINGAN SARAF TIRUAN UNTUK PREDIKSI DIABETES MENGGUNAKAN DATASET PIMA INDIANS

Feri Kusuma Wardhana¹, Nur Sumaya Jati², Bintang Radityo Seto³, Febry Mustika Wati⁴, Yesika Ayu Yuliana⁵, Rajnaparamitha Kusumastuti⁶

¹²³⁴⁵ Prodi Informatika, STMIK Amikom Surakarta

¹²³⁴⁵ Sukoharjo Indonesia

Email: ¹feriwardhana41@gmail.com, ²nursumaya035@gmail.com,
³bintangradityo84@gmail.com, ⁴febrymustika23@gmail.com,
⁵yesikaayuyu@gmail.com, ⁶rajna@dosen.amikomsolo.ac.id

Abstract

The application of Backpropagation Neural Network (BPNN) for diabetes prediction using the Pima Indians dataset represents a significant advancement in developing effective diagnostic tools. This study preprocesses the dataset, which consists of 768 records, by addressing missing values and outliers through statistical techniques and scaling. The data is then split into training (80%) and testing sets (20%), followed by feature scaling using StandardScaler. A grid search with cross-validation is employed to optimize the hyperparameters of the Multilayer Perceptron (MLP) classifier. The optimized MLP classifier achieved a cross-validation score of 78.50%. On the test set, the classifier attained an accuracy of 83%, with a precision of 0.85 for the non-diabetes class and 0.79 for the diabetes class. The model's performance is evaluated using classification reports and a confusion matrix, highlighting the potential of BPNN in medical diagnostics. These findings contribute to ongoing research on applying machine learning in healthcare, emphasizing the importance of data preprocessing and hyperparameter tuning.

Keywords: *Backpropagation, Diabetes Prediction, Healthcare, Neural Network, Pima Indians Dataset*

Abstraksi

Penerapan Jaringan Syaraf Tiruan Backpropagation (BPNN) untuk prediksi diabetes menggunakan dataset Pima Indians merupakan langkah signifikan dalam pengembangan alat diagnostik yang efektif. Penelitian ini memproses dataset yang terdiri dari 768 data dengan mengatasi nilai yang hilang dan outlier melalui teknik statistik serta penskalaan. Data kemudian dibagi menjadi set pelatihan (80%) dan set pengujian (20%), diikuti penskalaan fitur menggunakan StandardScaler. Optimasi hyperparameter untuk klasifikasi Multilayer Perceptron (MLP) dilakukan menggunakan Grid Search dengan cross-validation. MLP classifier yang dioptimalkan mencapai skor validasi terbaik sebesar 78,50%. Pada set pengujian, classifier memperoleh akurasi 83%, dengan presisi 0,85 untuk kelas non-diabetes dan 0,79 untuk kelas diabetes. Kinerja dievaluasi melalui laporan klasifikasi dan confusion matrix, menunjukkan potensi BPNN dalam diagnosis medis. Temuan ini mendukung penelitian dalam penerapan machine learning di bidang kesehatan, menekankan pentingnya pemrosesan data dan tuning hyperparameter.

Kata Kunci: *Backpropagation, Dataset Pima Indians, Jaringan Syaraf Tiruan, Kesehatan, Prediksi Diabetes*

1. PENDAHULUAN

Diabetes mellitus adalah sebuah penyakit yang kronis dimana diderita oleh sebuah masyarakat di seluruh penjuru dunia. Pada penyakit ini dapat ditandai oleh tingginya sebuah kadar gula dari darah yang dapat disebabkan oleh sebuah gangguan pada sebuah produksi dan sebuah fungsi insulin[1]. Prediksi dari dini terhadap sebuah diabetes akan sangat penting untuk mengurangi sebuah resiko dari komplikasi dan dapat meningkatkan kualitas dari hidup sebuah pasien. Oleh sebab itu. Adanya sebuah pengembangan dari model prediksi yang dapat akurat dan efisien menjadi sebuah focus banyak penelitian dari dalam beberapa dekade terakhir[2]

Pada penelitian ini memiliki tujuan untuk menerapkan sebuah metode dari backpropagation Jaringan Saraf Tiruan (JST) untuk melakukan prediksi diabetes dan menggunakan dataset Pima Indians. Pada metode JST dipilih karena sebuah kemampuannya dalam menangani sebuah data yang sangat kompleks dan menghasilkan sebuah prediksi yang cukup akurat. Backpropagation merupakan sebuah algoritma pada pelatihan JST yang sangat populer karena disebabkan oleh keefektifannya dalam meminimalkan sebuah kesalahan dari prediksi melalui sebuah penyesuaian dari bobot jaringan[3].

Dataset Pima Indians adalah sebuah dataset yang dipergunakan dalam sebuah penelitian ini berasal dari sebuah Institut Nasional yaitu NIDDK. Dataset ini berisi sebuah data medis dari wanita dari sebuah suku Indian Pima yang memiliki diatas dari 21 tahun. Data tersebut mencakup berbagai sebuah variable yang relevan dengan sebuah diabetes, seperti sebuah kadar glukosa, sebuah tekanan darah, sebuah indeks massa tubuh (BMI), dan sebuah riwayat keturunan . Penggunaan dari dataset ini sangat umum dalam sebuah penelitian berkaitan dengan prediksi diabetes. Karena pada dataset ini menawarkan sebuah informasi yang kaya dan juga mendetail yang juga dapat digunakan untuk membangun sebuah model dari prediksi yang lebih kuat.[4]

Dalam beberapa sebuah penelitian sebelumnya telah dapat menunjukkan potensi dari JST dalam sebuah prediksi diabetes. Namun , ada beberapa tantangan masih ada, yaitu termasuk sebuah overfitting dan sebuah kebutuhan akan sebuah data yang lebih besar untuk sebuah pelatihan dari model yang lebih efektif[3].

Dalam sebuah penelitian ini , penulis akan membahas sebuah metode dari yang digunakan untuk melakukan sebuah preprocessing data, arsitektur dari JST yang akan diimplementasikan, serta sebuah prosedur pelatihan dan evaluasi dari model. Penulis juga akan membandingkan dari hasil yang diperoleh dengan sebuah penelitian terdahulu untuk menilai sebuah keefektifan pendekatan yang digunakan.[5]

Dengan demikian, diharapkan pada sebuah penelitian yang penulis lakukan memberikan sebuah kontribusi dalam sebuah penelitian dari model prediksi diabetes ini serta juga memberikan sebuah wawasan baru bagi sebuah penelitian di masa depan. Melalui penerapan metode JST ini dengan sebuah algoritma backpropagation, penelitian ini berupaya untuk meningkatkan akurasi diabetes dan memberikan alat yang lebih efektif bagi preaktisi Kesehatan dalam mengidentifikasi resiko diabetes secara dini.

2. TINJAUAN PUSTAKA

Tinjauan pustaka ini membahas berbagai pendekatan machine learning yang telah diaplikasikan untuk memprediksi diabetes, dengan penekanan pada penerapan algoritma Backpropagation dalam jaringan syaraf tiruan (JST) dan penggunaan dataset Pima Indians.

Penelitian [6] melaporkan bahwa kombinasi metode SVM dengan kernel RBF serta penerapan teknik Forward Selection mampu memberikan hasil yang sangat baik. Metode ini menunjukkan performa akurasi sebesar 91,2%, precision 93,0%, recall 94,3%, dan F1-score 93,7%, dengan keunggulan dalam mendeteksi pola melalui optimalisasi kernel. Selain itu, studi [7] yang memanfaatkan algoritma Evolving Artificial Neural Network (EANN) mencatat tingkat akurasi 83,55%, meskipun performa model sangat bergantung pada kualitas dataset yang digunakan.

Studi [9] mengungkapkan bahwa penerapan Naïve Bayes tanpa metode ensemble seperti Adaboost atau Bagging menghasilkan tingkat presisi tertinggi sebesar 80,23%. Sementara itu, model Logistic Regression dengan tuning menggunakan Grid Search mencatatkan rata-rata akurasi sebesar 79%, dengan pengujian data tambahan mencapai akurasi hingga 83,33%.

Pada penelitian [10], algoritma Backpropagation berhasil diterapkan untuk prediksi diabetes, dengan akurasi mencapai 81,8%. Parameter yang digunakan meliputi 500 kali training cycle, learning rate sebesar 0,3, serta momentum 0,2. Rendahnya nilai Root Mean Square Error (RMSE), yaitu 0,3831, menunjukkan bahwa metode ini memiliki tingkat keandalan yang cukup tinggi.

Hasil pengujian pada studi [11] menggunakan validasi 10 kali lipat menunjukkan bahwa algoritma Decision Tree menghasilkan akurasi terbaik, yaitu 95,58%, dibandingkan dengan algoritma Naïve Bayes. Hal ini memperlihatkan bahwa Decision Tree unggul dalam pengolahan dataset diabetes.

Penelitian [12] memanfaatkan metode SVM dengan kernel Linear dan Polynomial degree 1 untuk menganalisis risiko pengidap diabetes. Rata-rata tingkat akurasi yang dicapai adalah 93%, dengan nilai tertinggi mencapai 96% dan terendah 92%. Sebagai perbandingan, algoritma K-Nearest Neighbor (KNN) memiliki rata-rata akurasi sebesar 79%, dengan hasil terbaik mencapai 100% dan terendah 37%.

Dalam penelitian [13], evaluasi performa algoritma dilakukan berdasarkan nilai presisi dan recall. Untuk kelas negatif, recall tercatat sebesar 99,06% dan presisi sebesar 97,54%. Sementara itu, untuk kelas positif, recall mencapai 96,00% dan presisi sebesar 98,46%.

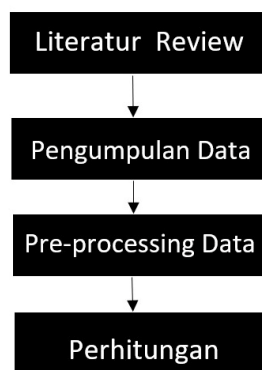
Penelitian [15] menunjukkan bahwa algoritma KNN juga dapat diterapkan dalam prediksi diabetes dengan tingkat akurasi sebesar 72%. Pengujian Blackbox yang dilakukan terhadap aplikasi berbasis algoritma ini memastikan bahwa semua fitur dapat berfungsi dengan baik, dengan tingkat keberhasilan pengujian mencapai 100%.

Secara keseluruhan, temuan dari berbagai penelitian ini menegaskan bahwa algoritma machine learning seperti SVM, Backpropagation, Naïve Bayes, dan KNN

memiliki potensi besar untuk prediksi diabetes. Penerapan preprocessing data yang tepat dan optimasi parameterisasi dapat lebih meningkatkan kinerja metode-metode tersebut.

3. METODE PENELITIAN

Metode penelitian ini, memuat sebuah tahapan-tahapan yang perlu dilakukan untuk membangun system prediksi penyakit diabetes ini. Di mulai dari literatur riew, pengumpulan data ,pre-processing data hingga perhitungan yang dapat dilihat pada gambar 1 dibawah ini.



Gambar 1. Tahap Penelitian

3.1. Pengumpulan Data

Pada dataset yang penulis gunakan dalam penelitian penulis dari National Institute of Diabetes and Digestive and Kidney Diseases. Pada Dataset ini berisi berbagai variable yang relevan dengan diabetes, seperti kadar glukosa, tekanan darah , indeks massa tubuh (BMI), dan riwayat keturunan . Berikut tersaji pada gambar 2 dari dataset penulis .

Row No.	DiabetesPe...	Pregnancies	Glucose	BloodPress...	SkinThickne...	Insulin	BMI	Age	Outc
1	0.627	6	148	72	35	0	33.600	50	1
2	0.351	1	85	66	29	0	26.600	31	0
3	0.672	8	183	64	0	0	23.300	32	1
4	0.167	1	89	66	23	94	28.100	21	0
5	2.288	0	137	40	35	168	43.100	33	1
6	0.201	5	116	74	0	0	25.600	30	0
7	0.248	3	78	50	32	88	31	26	1
8	0.134	10	115	0	0	0	35.300	29	0
9	0.158	2	197	70	45	543	30.500	53	1
10	0.232	8	125	96	0	0	0	54	1
11	0.191	4	110	92	0	0	37.600	30	0
12	0.537	10	168	74	0	0	38	34	1
13	1.441	10	139	80	0	0	27.100	57	0

Gambar 2. Dataset Pima Indians

3.2. Pre-Processing Data

Tahap ini mencakup beberapa langkah dasar untuk memastikan data dalam format yang optimal untuk pelatihan[6]. Pada dataset dapat ditemukan beberapa data yang memiliki missing value, noise, dan tidak konsisten. Sehingga tahapan preprocessing data dalam jaringan syaraf tiruan (JST) . backpropagation adalah salah satu dari tahapan penting sebelum data diolah oleh model[7]. Langkah pre-processing meliputi[8] :

- Mengganti sebuah nilai yang hilang dengan sebuah nilai median dari atribut yang bisa bersangkutan[9]
- mengubah skala pada data ke rentang dari [0,1] menggunakan sebuah metode min-max normalisasi untuk dapat memastikan setiap sebuah fitur memiliki kontribusi yang setara dalam sebuah proses training[10].

3.3. Perhitungan Backpropagation

Pada algoritma pelatihan backpropagation tersebut adalah [7]:

Tahap 0 : pada Inisialisasi pada sebuah nilai pada sebuah bobot dengan menggunakan sebuah nilai dari acak yang kecil[7].

Tahap 1 : Jika sebuah kondisi tersebut berhenti dan juga masih juga belum bisa terpenuhi, maka akan dikerjakan dari *tahap 2* sampai 9.[7]

Tahap 2 : pada setiap pasangan dari sebuah pelatihan (s : t), lanjutkan ke tahap 3 sampai 8.[7]

Tahap 3 : kemudian Setiap *input yang* ($X_i, i = 1, \dots, n$) menerima sebuah sinyal dari *input* x_i dan mendistribusikannya ke unit di sebuah lapisan selanjutnya (*hidden layer*).[7]

Tahap 4 : Dalam setiap sebuah unit pada lapisan tersebut ($Z_j, j = 1, \dots, p$) menghitung dari sebuah nilai input berdasarkan sebuah bobot yang diberikan :

$$z_{inj} = v_0j + \sum_{i=1}^n x_i v_{ij} \dots\dots\dots(1)$$

Selanjutnya akan menghitung sebuah nilai dari output itu. berdasarkan dari fungsi pada aktivasi akan dipilih [7]:

$$z_j = f(z_{inj}) \dots\dots\dots (2)$$

Hasilnya pada fungsi itu dapat didistribusikan ke seluruh unit pada sebuah *lapisan* berikutnya (*output layer*).

Tahap 5 : Untuk setiap sebuah unit *output* ($Y_k, k = 1, \dots, m$), hitung sebuah nilai pada input menggunakan bobot yang sesuai [7]:

$$y_{ink} = w_0k + \sum_{j=1}^p z_j w_{jk} \dots\dots\dots (3)$$

Selanjutnya, nilai output dihitung berdasarkan fungsi aktivasi :

$$y_k = f(y_{ink}) \dots\dots\dots (4)$$

Tahap 6 : Untuk Setiap sebuah unit output ($Y_k, k = 1, \dots, m$) Menerima dari target sebuah pola yang sesuai dengan sebuah input pola dan kemudian menghitung sebuah error pada informasi [7]:

$$\delta_k = (t_k - y_k) f'(y_{in_k}) \dots\dots\dots (5)$$

selanjutnya, lakukan penyesuaian dari bobot untuk memperbaiki dari nilai bobot[7] :

$$w_{jk} : \Delta w_{jk} = \alpha \delta_k z_j \dots\dots\dots (6)$$

Hitung sebuah penyesuaian bias [7] :

$$w_{0k} : \Delta w_{0k} = \alpha \delta_k \dots\dots\dots (7)$$

dan nilai δ_k dikirim ke sebuah unit pada lapisan sebelumnya.

Tahap 7 : Setiap unit di lapisan tersembunyi ($Z_j, j = 1, \dots, p$), menghitung sebuah delta berdasarkan dari input ddari unit di lapisan atasnya[7]:

$$\delta_{in_j} = \sum^m \delta_k k = w_j \dots\dots\dots (8)$$

Selanjutnya, nilai ini akan dikalikan dengan sebuah turunan dari fungsi aktivasi tersebut untuk menghitung erorr informasi[7]:

$$\delta_j = (\delta_{in_j}) f'(z_{in_j}) \dots\dots\dots (9)$$

Hitung koreksi dari bobot untuk memperbaiki nilai dari u_{ij} [7] :

$$\Delta u_{ij} = \alpha \delta_j x_i \dots\dots\dots (10)$$

dan juga koreksi bias [7] :

$$v_{oj} \Delta v = \alpha \delta \dots\dots\dots (11)$$

Tahap 8 : Pada setiap unit output ($Y_k, k=1, \dots, m$) memperbaiki bias dan bobotnya [7] :

$$w_{jk} (baru) = w_{jk} (lama) + \Delta w_{jk} \dots\dots\dots (12)$$

Setiap sebuah unit hidden ($Z_j, j=1, \dots, p$) update bias dan sebuah bobotnya ($i=0, \dots, n$) [7]:

$$v_{ij} (baru) = v_{ij} (lama) + \Delta v_{ij} \dots\dots\dots (13)$$

Langkah - langkah pada backpropagation ini akan menggambarkan sebuah proses dari sebuah training suatu jaringan syaraf tiruan dengan menggunakan backpropagation yang meliputi sebuah inisialisasi bobot kemudian feedforward kemudian backpropagation erorr dan juga update dari bobot hingga kondisi tersebut berhenti terpenuhi[11]

Metode penelitian ini juga sudah mencakup dari tahapan penting dalam sebuah pre-processing data dan juga perhitungan backpropagation yang akan memastikan dari sebuah model yang dapat dilatih dengan menggunakan data yang bersih dan optimal, serta mengupdate bobot juga berdasarkan error yang dihasilkan dari selama proses pelatihan tersebut

3.4. Pengujian Model

Setelah pelatihan model JST selesai, tahap pengujian dilakukan guna menilai efektivitas kinerjanya. Dataset dibagi menjadi dua bagian: satu untuk pelatihan dan yang lain untuk pengujian. Model dioptimalkan melalui proses Grid Search dengan cross-validation, yang bertujuan untuk menemukan kombinasi terbaik dari hyperparameter pada model MLPClassifier. Evaluasi hasil pengujian dilakukan menggunakan beberapa metrik berikut:

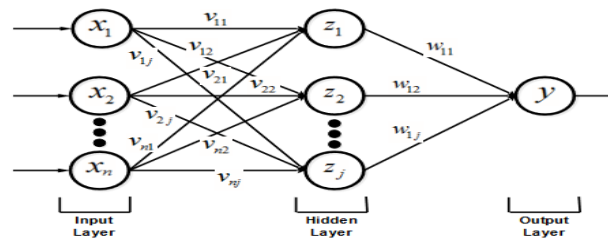
- a) Akurasi, Mengukur persentase prediksi yang benar pada data pengujian.
- b) Presisi, Menggambarkan ketepatan model dalam mengklasifikasikan kelas diabetes dan non-diabetes secara akurat.
- c) Confusion Matrix, Matriks ini menggambarkan kemampuan model dalam membedakan antara kelas diabetes dan non-diabetes.

Dari hasil pengujian, model menunjukkan akurasi sebesar 83%, dengan presisi 0.85 untuk kelas non-diabetes dan 0.79 untuk kelas diabetes. Hal ini menunjukkan potensi yang menjanjikan dari BPNN dalam mendukung diagnosis di bidang medis.

3.5. Backpropagation JST

JST adalah sebuah model dari neural network yang berbasis pada multilayer perceptron. multilayer perceptron adalah sebuah tipe jaringan syaraf tiruan yang mempunyai banyak sekali lapisan, termasuk pada satu atau lebih lapisan yang tersembunyi yang berada di antara dari lapisan input dan juga output.[12]

JST adalah sebuah sistem dari berbasis neuron bisa meniru sebuah cara kerja dari otak manusia untuk melakukan proses sebuah informasi dan sebuah mengenali dari pola yang kompleks. Pada gambar dibawah menunjukkan sebuah struktur dari multilayer perceptron dengan sebuah n buah unit input (x_1, x_2, \dots, x_n), sebuah satu lapisan tersembunyi yang memiliki sebuah m buah unit (z_1, z_2, \dots, z_m) dan sebuah unit dari output tunggal [13]. Arsitektur multiplayer perception dapat dilihat pada gambar 4.



Gambar 4. Arsitektur Multiplayer Perceptron

Maksud dari Backpropagation adalah sebuah algoritma pelatihan yang diawasi yang umumnya digunakan pada sebuah multilayer perceptron untuk menyesuaikan bobot – bobot yang menghubungkan neuron-neuron di lapisan tersembunyi . proses pembelajaran dilakukan dengan memasukkan sebuah data ke jaringan. Kemudian informasi dari lapisan pada input akan diteruskan ke lapisan tersembunyi untuk dilakukan proses , dan hasil prediksinya dari lapisan output tersebut dibandingkan dengan nilai yang diharapkan untuk menghitung kesalahan [14].

4. HASIL DAN PEMBAHASAN

4.1. Preprocessing

Pra-pemrosesan data adalah sebuah langkah dimana untuk memastikan dalam sebuah kondisi yang sangat optimal sebelum digunakan dalam sebuah pelatihan model[15]. Dalam penelitian ini , dimana dataset dipergunakan adalah Pima Indian Diabetes Dataset , dimana dataset ini merupakan dataset yang memiliki beberapa sebuah fitur dan sebuah label dari kelas (“Outcome”). Untuk langkah - Langkah nya adalah pra-pemrosesan yang dapat dilakukan adalah sebagai berikut [8]:

1. Mengatasi sebuah nilai – nilai yang hilang atau tidak benar dalam sebuah dataset.
2. Memisahkan sebuah dataset menjadi sebuah subset suatu pelatihan dan juga dalam suatu pengujian.
3. Menggunakan sebuah standardscaler dimana untuk melakukan menormalkan fitur – fitur pada dataset tersebut .

4.1.1. Pembersihan Data

Pada dataset awal memiliki sebuah beberapa sebuah nilai nol yang ada pada pada fitur – fitur yang seharusnya tidak mungkin bernilai nol (misalnya sebuah tekanan darah, kadar glukosa). Oleh karena itu pada sebuah nilai-nilai nol ini dimana digantikan dengan sebuah median dari sebuah masing -masing fitur tersebut.

4.1.2. Pembagian data

Dataset ini dibagi menjadi sebuah subset dari pelatihan dan sebuah pengujian dengan proporsi 80:20. Subset data pelatihan dapat digunakan untuk melatih sebuah model, sedangkan subset data pengujian dipakai untuk menilai kinerja dari sebuah model.

4.1.3. Fitur – fitur dataset dinormalisasi

menggunakan sebuah StandardScaler untuk memastikan bahwa setiap fitur memiliki sebuah mean 0 dan juga dimana standard deviation berupa 1. Hal ini juga penting untuk sebuah algoritma pembelajaran mesin seperti MLPClassifier yang sensitive terhadap skala data.

4.2. Analisis Data

4.2.1. Hasil Skenario 1

Pada hasil Penelitian ini pada pengujian 70 banding 30 dengan MLPClassifier yang diterapkan untuk memprediksi sebuah diabetes menggunakan sebuah dataset Pima Indians tersebut. Pada model ini dilatih dan diuji menggunakan sebuah Teknik cross-validation dengan menggunakan 3 fold dan sebuah parameter turning melalui sebuah GridSearchCV. Pada proses GridSearchCV mengevaluasi total 23328 fitting dengan 7776 kandidat parameter. Berikut adalah sebuah hasil yang diperoleh dari eksperimen ini :

- Best Parameters:
Aktivasi : relu , Alpha : 0.0001 , Beta 1 : 0.9 , Beta 2 : 0.999 , Ukuran Lapisan Tersembunyi : Dua lapisan dengan masing – masing 100 neuron , Learning Rate : Konstan , Inisialisasi Learning Rate : 0.01 , Iterasi Maksimal : 200 dan Solver : Stochastic Gradient Descent (SGD)
- Best Cross-Validation Score: 0.7597765363128492

Classification Report on Test Set:				
	precision	recall	f1-score	support
0	0.79	0.87	0.83	150
1	0.71	0.58	0.64	81
accuracy			0.77	231
macro avg	0.75	0.73	0.74	231
weighted avg	0.77	0.77	0.76	231

Gambar 5. Hasil analisis 70:30

Berdasarkan hasil pada gambar 5, hasil eksperimen tersebut menunjukkan bahwa model MLPClassifier dengan sebuah parameter yang dioptimalkan dengan mencapai akurasi 77 % pada dataset uji. Tabel dari klasifikasi menunjukkan bahwa model ini memiliki sebuah precision dan recall yang baik untuk kelas 0 (non-diabetes), namun pada performa untuk kelas 1 (diabetes) lebih rendah.

4.2.2. Hasil Skenario 2

Pada pengujian ini menggunakan perbandingan data pelatihan 80% banding 20% untuk data pengujian, sebuah mode MLPClassifier tersebut diterapkan untuk

memprediksi sebuah diabetes dengan menggunakan dataset pima Indians. Model ini dilatih dan diuji menggunakan Teknik cross-validation dengan 3 fold dan parameter turning melalui GridSearchCV. Pada proses GridSearchCV mengevaluasi dari total 23328 dengan fitting dengan 7776 kadidat parameter . Berikut adalah sebuah hasil yang diperoleh dari eksperimen tersebut :

A. Best Parameters :

Fungsi Aktivasi : Tanh , Alpha : 0.0001 , Beta 1 : 0.99 , Beta 2 : 0.99 , Ukuran Lapisan Tersembunyi : Tiga lapisan dengan ukuran masing – masing 50 , 100 , dan 50 neuron , Tingkat Pembelajaran : Konstan , Inialisasi Tingkat Pembelajaran : 0.001 , Interasi Maksimal : 200 , Metode Solver : Adam

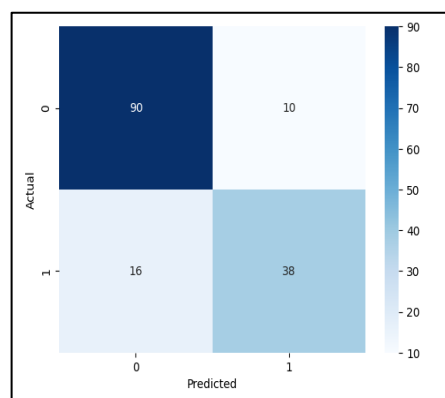
B. Best Cross-Validation Score : 0.7849912322652638

Setelah menemukan parameter terbaik, model dilatih kembali dengan parameter tersebut dan dievaluasi pada dataset uji. Berikut laporan klasifikasi untuk set uji.

C. Classification Report: Hasil pengujian 80:20 dan hasil confusion matrix dapat dilihat pada gambar 6 dan 7.

Classification Report on Test Set:				
	precision	recall	f1-score	support
0	0.85	0.90	0.87	100
1	0.79	0.70	0.75	54
accuracy			0.83	154
macro avg	0.82	0.80	0.81	154
weighted avg	0.83	0.83	0.83	154

Gambar 6. Hasil Pengujian 80 : 20



Gambar 7. Confusion Matrix

Hasil pada eksperimen dengan bandingan 80 : 20 menunjukkan bahwa sebuah model MLPClassifier dengan parameter yang telah dioptimalkan mampu mencapai sebuah akurasi 83% pada dataset uji. Pada tabel klasifikasi menunjukkan bahwa sebuah model ini memiliki sebuah precision dan recall yang baik untuk kelas 0 (non-diabetes), namun untuk performa untuk kelas 1 (diabetes) sedikit lebih rendah.

Beberapa poin penting dari hasil ini adalah sebagai berikut adalah :

- parameter terbaik yang ditemukan melalui sebuah GridSearchCV menunjukkan bahwa penggunaan dari fungsi aktivasi 'tanh' dan solver 'adam' dengan melakukan konfigurasi lapisan tersembunyi (50,100,50), Learning rate yang konstan dengan nilai alpha yang rendah (0.0001) dan beta yang tinggi (0.99) juga memberikan sebuah kontribusi positif terhadap sebuah model
- Model ini menunjukkan bahwa performa yang lebih baik dalam mengklasifikasikan sebuah non-diabetes (kelas 0) dibandingkan dengan sebuah diabetes (kelas 1) . hal ini mungkin juga disebabkan oleh distribusi data yang tidak seimbang dalam sebuah dataset, di mana jumlah dari sampel non-diabetes lebih banyak dibandingkan sampel dari diabetes tersebut.
- Skor cross-validation yang mencapai hamper 78.5% menunjukkan bahwa sebuah model memiliki generalisasi yang cukup baik terhadap data yang tidak terlihat selama dalam pelatihan, meskipun terdapat sedikit dari penurunan performa pada sebuah dataset uji
- Proses tuning hyperparameter dengan menggunakan sebuah GridSearchCV memungkinkan eksplorasi dari berbagai kombinasi parameter untuk menemukan sebuah konfigurasi yang optimal , meskipun memerlukan sebuah waktu dan sumber daya komputasi yang signifikan.

Secara Keseluruhan , model MLP Classifier yang telah dioptimalkan ini menunjukkan performa yang cukup baik dalam memprediksi diabetes pada dataset Pima Indians. Namun, untuk meningkatkan performa lebih lanjut, beberapa Langkah seperti peningkatan iterasi maksimum, penanganan ketidakseimbangan data, atau eksplorasi lebih lanjut terhadap Teknik preprocessing dan arsitektur jaringan dapat dipertimbangkan. Berikut ini perbandingan analisis data tersaji pada tabel 1.

Tabel 1. Perbandingan analisis data

Nama	Method	Dataset	Akurasi
Muhammad Salsabil, Nuril Lutvi Azizah, Ade Eviyanti (2024)	Random Forest dan XGBoost	Pima Indians	74 %
Analisis penulis 70 : 30	Backpropagation JST	Pima Indians	77 %
Analisis penulis 80 : 20	Backpropagation JST	Pima Indians	83%

Dilihat dari beberapa analisis sebelumnya dan analisis dari beberapa jurnal terkait terlihat bahwa analisis terbaik adalah analisis dengan method Backpropagation dengan perbandingan data raining dan data testing adalah 80 :20 dengan mode MLPClassifier tersebut diterapkan untuk memprediksi sebuah diabetes dengan menggunakan dataset pima Indians. Model ini dilatih dan diuji menggunakan Teknik cross-validation dengan 3 fold dan parameter turning melalui GridSearchCV. Pada proses GridSearchCV mengevaluasi dari total 23328 dengan fitting dengan 7776 kadidat parameter. Akan tetapi masih bisa dilakukan penelitian lebih lanjut untuk meningkatkan akurasi menjadi lebih

baik dengan menambahkan sebuah SMOTE atau sebuah penyeimbang perhitungan pada dataset.

5. KESIMPULAN

Dari hasil penelitian tersebut, penulis mendapatkan kesimpulan bahwa model MLPClassifier dengan sebuah optimasi hyperparameter dan GridSearchCV dapat digunakan untuk prediksi diabetes pada dataset Pima Indians dengan akurasi sebesar 83% pada dataset testing. Model memiliki performa yang baik dalam mengklasifikasikan pasien yang tidak menderita diabetes, namun masih ada ruang untuk perbaikan dalam mengklasifikasikan pasien yang menderita diabetes. Beberapa Langkah yang dapat diambil untuk peningkatan performa model di masa depan termasuk :

1. Meningkatkan jumlah iterasi pada proses pelatihan untuk mencapai konvergensi yang optimal. dan Menerapkan Teknik penyeimbangan data, seperti SMOTE, untuk mengatasi masalah ketidak seimbangan kelas.
2. Melakukan turning lebih lanjut terhadap parameter dan mencoba arsitektur yang berbeda.

DAFTAR PUSTAKA

- [1] R. Rizki, R. Athallah, I. Cholissodin, and P. P. Adikara, "Prediksi Potensi Pengidap Penyakit Diabetes berdasarkan Faktor Risiko Menggunakan Algoritme Kernel K-Nearest Neighbor," 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [2] "Implementasi Data Mining Dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode Random Forest Dan Xgboost," *Jurnal Ilmiah Komputasi*, vol. 23, no. 1, Mar. 2024, doi: 10.32409/jikstik.23.1.3507.
- [3] C. I. Damanik, D. Haratama, H. S. Tambunan, K. Kunci, J. ; Backpropogation, and M. Baru, "BULLETIN OF COMPUTER SCIENCE RESEARCH Penerapan JST Backpropagation Untuk Memprediksi Data Penerimaan Mahasiswa Baru Pada Universitas Simalungun," *Media Online*, vol. 2, no. 1, p. 33, 2021, [Online]. Available: <https://hostjournals.com/bulletincsr>

- [4] A. Nur Rais, E. Rahmawati, and R. Faizal Amir, "Prediksi Pima Indians Diabetes Database Dengan Ensemble Adaboost Dan Bagging," *Jurnal Sains dan Manajemen*, vol. 9, no. 2, 2021, [Online]. Available: <https://www.kaggle.com/uciml/pima->
- [5] Y. A. Hambali, R. Megasari, and R. R. Santoso, "Implementasi Metode Machine Learning menggunakan Algoritma Evolving Artificial Neural Network pada Kasus Prediksi Diagnosis Diabetes Implementation of Machine Learning Method using Evolving Artificial Neural Network Algorithm in Diabetes Diagnosis Prediction Case," 2023. [Online]. Available: <https://ejournal.upi.edu/index.php/JATIKOM>
- [6] M. Sholeh, D. Andayati, R. Yuliana Rachmawati, P. Studi Informatika, and F. Teknologi Informasi dan Bisnis, "DATA MINING MODEL KLASIFIKASI MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DENGAN NORMALISASI UNTUK PREDIKSI PENYAKIT DIABETES DATA MINING MODEL CLASSIFICATION USING ALGORITHM K-NEAREST NEIGHBOR WITH NORMALIZATION FOR DIABETES PREDICTION."
- [7] D. L. Rahakbauw, F. Tahya, and M. Sopaheluwakan, "Prosiding Seminar Nasional Pendidikan Matematika Universitas Pattimura | Agustus," 2019.
- [8] F. N. Fajri, "Prediksi Penyakit Diabetes dengan Menggunakan Artificial Neural Network."
- [9] E. Hasmin, C. Susanto, K. Aryasa, U. Dipa Makassar, and J. Perintis Kemerdekaan Km, "Sistem Pakar Prediksi Penyakit Diabetes Menggunakan Metode K-NN Berbasis Android Expert System for Predicting Diabetes Using the Android-Based K-NN Method," *Cogito Smart Journal |*, vol. 8, no. 2.
- [10] P. Fingki Marwati and R. Fauzi, "PENYAKIT DIABETES MELITUS MENGGUNAKAN JARINGAN SYARAF TIRUAN DENGAN METODE BACKPROPAGATION," pp. 26–34, doi: 10.55903/jitu.v2i1.163.
- [11] A. F. Suahati, A. A. Nurrahman, and O. Rukmana, "Penggunaan Jaringan Syaraf Tiruan – Backpropagation dalam Memprediksi Jumlah Mahasiswa Baru," *Jurnal Media Teknik dan Sistem Industri*, vol. 6, no. 1, p. 21, Mar. 2022, doi: 10.35194/jmtsi.v6i1.1589.
- [12] A. Abdillah and I. Artikel, "PENERAPAN ALGORITMA BAYESIAN REGULARIZATION BACKPROPAGATION UNTUK MEMPREDIKSI PENYAKIT DIABETES," *Jurnal MIPA*, vol.

- 39, no. 2, pp. 150–158, 2016, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/JM>
- [13] L. G. Wright *et al.*, “Deep physical neural networks trained with backpropagation,” *Nature*, vol. 601, no. 7894, pp. 549–555, Jan. 2022, doi: 10.1038/s41586-021-04223-6.
- [14] A. Yuberta, “Jaringan Syaraf Tiruan dengan Algoritma Backpropagation dalam Memprediksi Hasil Asesmen Nasional Berbasis Komputer (ANBK) SMP Se Kota Sawahlunto,” *Jurnal Informasi dan Teknologi*, pp. 200–205, Sep. 2022, doi: 10.37034/jidt.v4i4.234.
- [15] R. Rizki, R. Athallah, I. Cholissodin, and P. P. Adikara, “Prediksi Potensi Pengidap Penyakit Diabetes berdasarkan Faktor Risiko Menggunakan Algoritme Kernel K-Nearest Neighbor,” 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>