

PERBANDINGAN ALGORITMA NEURAL NETWORK DENGAN LINIER DISCRIMINANT ANALYSIS (LDA) PADA KLASIFIKASI PENYAKIT DIABETES

Fida Maisa Hana, Deka Setia Negara, Khoirul Umam Khaqiqi

fidamaisa@umkudus.ac.id

Universitas Muhammadiyah Kudus

Abstrak

Salah satu penyakit kronis yang banyak diderita oleh penduduk Indonesia adalah Diabetes Melitus (DM), penyakit ini ditandai dengan nilai kadar glukosa dalam darah di atas normal. Penyakit ini termasuk penyakit yang rumit dan mematikan, oleh karena itu dibutuhkan perawatan medis yang kontinu agar resiko terjadinya komplikasi bisa dihindari. Guna menganalisa pasien pengidap penyakit diabetes sejak dini, Pencatatan terhadap penyakit ini banyak dilakukan agar dapat dilakukan pencegahan. Salah satu yang dilakukan adalah dengan menggunakan teknik klasifikasi data mining. Teknik klasifikasi digunakan untuk memprediksi pasien mana yang terkena penyakit diabetes dan tidak. Dalam penelitian ini menggunakan Algoritma kasifikasi data mining neural network dan linier Discriminant Analysis (LDA). Hasil penelitian menunjukkan akurasi sebesar 90.38% dengan algoritma Linear Discriminant Analysis (LDA) dan akurasi sebesar 95,19% didapat pada saat menggunakan algoritma Neural Network. Algoritma Neural Network menghasilkan akurasi lebih baik daripada algoritma Linear Discriminant Analysis (LDA) dalam klasifikasi penyakit diabetes.

Kata Kunci: Data Mining, Klasifikasi, Neural Network, LDA, Diabetes

Abstract

One of the chronic diseases that many Indonesians suffer from is Diabetes Mellitus (DM), this disease is characterized by the value of glucose levels in the blood above normal. This disease is a complex and deadly disease, therefore it requires continuous medical care so that the risk of complications can be avoided. In order to analyze patients with diabetes from an early age, many records of this disease can be done so that prevention can be done. One thing that is done is by using data mining classification techniques. Classification techniques are used to predict which patients will develop diabetes and which do not. In this study, the data mining neural network classification algorithm and linear Discriminant Analysis (LDA) were used. The results showed an accuracy of 90.38% with the Linear Discriminant Analysis (LDA) algorithm and an accuracy of 95.19% obtained when using the Neural Network algorithm. The Neural Network algorithm provides better accuracy than the Linear Discriminant Analysis (LDA) algorithm in the classification of diabetes.

Keywords: Data Mining, Classification, Neural Network, LDA, Diabetes

I. PENDAHULUAN

Diabetes mellitus adalah penyakit metabolis yang kronis yang mana pasien penyakit diabetes tidak menghasilkan jumlah insulin yang cukup atau bisa dikatakan tubuh pasien tidak sanggup memanfaatkan insulin dengan baik sehingga menyebabkan gula darah di dalam tubuh mengalami jumlah yang berlebihan, kondisi ini sering kali dirasakan setelah komplikasi terjadi pada organ tubuh (Khairani, 2012). Pasien didiagnosa menderita penyakit diabetes pada saat kadar

glukosa darahnya melebihi nilai normal (Nurlina, 2019). Penyakit diabetes melitus adalah penyakit yang memiliki kompleksitas tinggi, perawatan medis yang berkelanjutan sangat dibutuhkan guna menurunkan dampak komplikasi dengan pengecekan glikemik (ADA, 2016).

Banyaknya penderita diabetes dari tahun ke tahun semakin bertambah. Pasien diabetes di Indonesia sebesar 10 juta jiwa di tahun 2015. Merujuk pada data Federasi Diabetes Internasional, diprediksi penderita penyakit diabetes di Indonesia akan bertambah menjadi

16.2 pada tahun 2040 (Suwarno and AA Abdillah, 2017). Guna menyikapi masalah ini, perlu adanya pendeteksian sejak dini penyakit diabetes. Deteksi sejak dini diharapkan dapat menurunkan resiko komplikasi pada pasien diabetes diwaktu mendatang. Guna menganalisa pasien pengidap penyakit diabetes sejak dini, Pencatatan terhadap penyakit ini banyak dilakukan agar dapat dilakukan pencegahan. Salah satu yang pencatatan yang bisa dilakukan adalah dengan memanfaatkan teknik klasifikasi pada data mining.

Data mining adalah proses menemukan korelasi, pola, dan tren baru yang bermakna dengan cara memilih-milih data dalam jumlah besar yang disimpan dalam repositori, memanfaatkan teknologi pengenalan pola serta teknik statistik dan matematika (Apostolakis, 2010). Berdasarkan tugasnya, data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering, Asosiasi (Larose and Larose, 2014).

Salah satu tugas utama dalam data mining adalah klasifikasi. Klasifikasi adalah teknik mengenali suatu karakteristik dari data dan mengelompokkannya ke dalam kelas-kelas. Proses klasifikasi adalah proses menghitung data yang ada sebelumnya atau disebut juga data training dengan data baru atau data testing. Proses ini akan menghasilkan kemungkinan dalam data testing. (Hana, 2020).

Data mining berpengaruh dalam berbagai bidang, salah satunya di dunia kesehatan berperan dalam deteksi penyakit. Oleh karena ini dalam penelitian ini akan memanfaatkan ilmu data mining khususnya klasifikasi penyakit diabetes. Penelitian ini akan membandingkan dua metode klasifikasi Neural Network dan linear discriminant analysis (LDA).

Neural Network adalah sebuah metode yang memiliki kemampuan untuk belajar dari beberapa contoh karena memiliki karakteristik yang dapat belajar dari data-data sebelumnya dan mengenal pola data yang selalu berubah. Neural Network menyerupai jaringan saraf pada manusia. Pada penelitian yang dilakukan oleh (Maulidi *et al.*, 2018) dengan judul “Penerapan Neural Network Backpropagation Untuk Klasifikasi Artikel

Clickbait” menghasilkan akurasi sebesar 85% dan pada penelitian yang dilakukan oleh (Hadianto, Novitasari and Rahmawati, 2019) dengan judul “Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network” menghasilkan akurasi sebesar 98,24 %. Hal ini menunjukkan bahwa menggunakan metode Neural Network bisa menghasilkan akurasi yang baik.

LDA adalah sebuah metode yang digunakan untuk pengenalan pola pada perhitungan statistika dengan cara menemukan proyeksi linear dari data yang bertujuan untuk memaksimumkan penyebaran antar kelas dan meminimumkan penyebaran dalam kelas data (Budiman, Santoso and Afrianto, 2017). Pada penelitian yang dilakukan oleh (Budiman, Santoso and Afrianto, 2017) dengan judul “Pendeteksi Jenis Autis pada Anak Usia Dini Menggunakan Metode Linear Discriminant Analysis (LDA)” menghasilkan akurasi sebesar 88%. Pada penelitian yang dilakukan oleh (Ramdhani, 2015) dengan judul “Komparasi Algoritma Lda Dan Naïve Bayes Dengan Optimasi Fitur Untuk Klasifikasi Citra Tunggal Pap Smear” menunjukkan bahwa algoritma LDA menghasilkan akurasi lebih tinggi dibandingkan dengan Naïve Bayes yaitu sebesar 98,87%. Dari kedua penelitian menunjukkan bahwa menggunakan metode LDA dalam klasifikasi menghasilkan nilai akurasi yang tinggi.

Merujuk pada penjelasan di atas maka penelitian ini akan melakukan perbandingan algoritma neural network dengan linier discriminant analysis (LDA) pada klasifikasi penyakit diabetes.

II. LANDASAN TEORI

A. Data Mining

Metode yang dipakai untuk menggali informasi yang belum ditemukan dengan cara manual dari suatu kumpulan data disebut dengan data mining (Dita Merawati, 2019). Sejak tahun 1990, sudah mulai dikenal istilah data mining hal ini karena kebutuhan mengolah data adalah hal yang bermanfaat dan sangat perlu dilakukan dari bermacam bidang ilmu, dalam bidang kesehatan, akademik, bahkan industri (Gorunescu, 2011). Ilmu Data mining adalah perpaduan ilmu dari

artificial intelligence, statistik, dan penelitian basis data yang selalu meningkat. Menurut artikel (Larose and Larose, 2014) metode data mining merupakan sebuah proses menentukan ikatan yang mengandung arti, pola, dan keterkaitan dengan mengolah kelompok data. Dalam data mining terdapat 6 metode yang biasa di jalankan yaitu ramalan atau prediksi, penggambaran atau deskripsi, klasifikasi, estimasi, asosiasi dan clustering (Larose and Larose, 2014).

a. Deskripsi

Tujuan dari operasi adalah untuk mengenal pola yang terbentuk berulang-ulang pada sekelompok data, selanjutnya menjadikan pola data tersebut membentuk kriteria dan aturan yang gampang dipahami.

b. Prediksi atau ramalan

Ramalan mempunyai kemiripan dengan teknik klasifikasi, tapi di sini data bagi kelas sesuai dengan perilaku atau nilai yang diprediksi pada waktu mendatang.

c. Klasifikasi

Klasifikasi atau pengelompokan adalah teknik mengenali suatu karakteristik dari data dan data tersebut dikelompokkan dalam suatu kelas-kelas.

d. Estimasi

Estimasi atau bisa disebut menerka, teknik ini memiliki kesamaan dengan prediksi, tapi di sini yang diestimasi berupa bilangan numerik.

e. Clustering

Clustering adalah klasterisasi atau pengelompokan data dengan kelas data yang mempunyai karakteristik sama dijadikan satu kelompok dan yang berbeda dikelompok yang lain.

f. Asosiasi

Teknik asosiasi adalah mencari atribut yang nampak pada kondisi tertentu atau menghasilkan aturan assosiatif antara suatu kombinasi item.

B. Klasifikasi

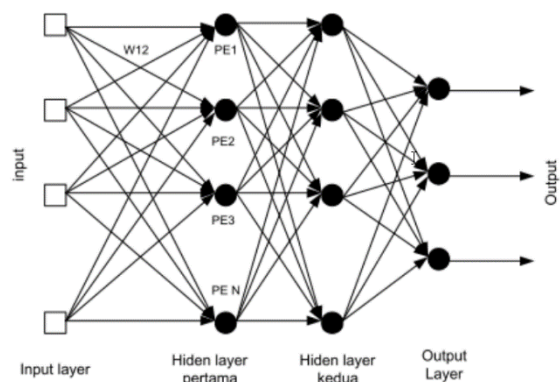
Proses mencari sebuah karakteristik data dan dipetakan dalam kelas-kelas sesuai dengan karakteristiknya masing-masing disebut dengan klasifikasi. Pada klasifikasi proses mencari karakteristik sebuah objek dilakukan, selanjutnya objek dengan karakteristik yang sama dimasukkan ke dalam

salah satu kelas yang sudah diartikan terlebih dahulu(Larose and Larose, 2014). Proses klasifikasi adalah proses menghitung data yang ada sebelumnya atau disebut juga data training dengan data baru atau data testing. Proses ini akan menghasilkan kemungkinan dalam data testing.

Dalam klasifikasi dataset yang digunakan harus memiliki label atau atribut tujuan. Meramal objek kelas pada setiap persoalan dalam data adalah tujuan dari klasifikasi. Dimulai dengan satu set data di mana kelas dikenal adalah sebuah tugas klasifikasi. Adapun Jenis masalah klasifikasi paling sederhana adalah klasifikasi biner (Putra and Chan, 2018).

C. Algoritma Neural Network

Neural Network (NN) merupakan metode dengan sistem kerja meniru fungsi otak manusia. Dalam otak manusia diyakini terdiri dari jutaan neuron atau unit pengolahan kecil yang sistem kerjanya paralel. Neuron saling terhubung satu sama lain melalui koneksi neuron (Rifai, 2013). Neural network dapat dianggap sebagai sistem kotak hitam yang menerima masukan dari lingkungan dan menghasilkan keluaran. Elemen pemrosesan dan pembobotan terkandung dalam neural network, elemen tersebut saling terhubung. Setiap susunan dalam jaringan berisi oleh kelompok elemen pemrosesan seperti yang ditunjukkan pada gambar.2 (Hadianto, Novitasari and Rahmawati, 2019)



Gambar 1. Multi Layer Perceptron Neural Networks

Setiap elemen pemrosesan mengumpulkan nilai dari semua input yang terhubung ke elemen pemrosesan dan menghasilkan output melalui operasi. Dari gambar di atas bisa dilihat bahwa ada tiga susunan yang

membangun neural network, yaitu *input layer*, *hidden layer*, dan *output layer*.

Artificial Neural Network (Jaringan Syaraf Tiruan) merupakan bentuk non-linear yang kompleks, dibentuk dari elemen yang secara individu mempunyai karakter mirip dengan model regresi. Jaringan syaraf tiruan dapat digambarkan dengan grafik, dan beberapa sub-grafik yang memiliki karakter yang sama dengan grafik sebelumnya.

Algoritma yang paling populer dalam pembelajaran jaringan saraf tiruan adalah algoritma back-propagation. Algoritma back-propagation merupakan algoritma yang memperkecil tingkat kesalahan dengan cara penyesuaian bobot berdasarkan perbedaan kelauran dan target yang akan dicapai. (Hadianto, Novitasari and Rahmawati, 2019). Algoritma Backpropagation bekerja dengan cara mundur, kerjanya dari *output layer* menuju *input layer* guna memperbaharui nilai pada *hidden layer* berdasarkan pada nilai kesalahan yang didapat. Adapun langkah-langkah cara kerja algoritma Backpropagation Neural Network adalah sebagai berikut:

1. Langkah pertama cara kerjanya adalah dimulai dengan *input layer*, hitung output dari setiap elemen pemroses melalui *input layer*.
2. Hitung error pada *output layer* yang merupakan selisih antara data aktual dan target.
3. Transformasikan error tersebut pada bagian yang sesuai di sisi input elemen pemroses.
4. Propagasi balik error ini pada output setiap elemen pemroses ke error yang terdapat pada input. Ulangi proses ini sampai input tercapai.
5. Ubah seluruh bobot dengan menggunakan kesalahan pada sisi elemen input dan elemen output pemroses yang terhubung. (Nugraha *et al.*, 2013)

Persamaan untuk menghitung neuron pada lapisan tersembunyi :

$$Y_in_j = V_{0j} + \sum_{i=1}^n X_i V_{ij} \quad (1)$$

$$Y_j = f(Y_in_j) \quad (2)$$

n = jumlah input

V = bobot

X = nilai input

Persamaan untuk menghitung neuron pada lapisan output :

$$Z_in_j = W_{0j} + \sum_{i=1}^n Y_i W_{ij} \quad (3)$$

$$Z_j = f(Z_in_j) \quad (4)$$

n = jumlah input

W = bobot

Y = nilai input

Persamaan untuk menghitung nilai error :

$$\delta_j = t_j - Z_j \quad (5)$$

t = nilai target

Z = nilai output

Persamaan untuk fungsi aktivasi :

$$f(x) = \frac{1}{1 + (e^{-x})} \quad (6)$$

$$f' = f(x)[1 - f(x)] \quad (7)$$

Persamaan untuk mengubah bobot :

$$\Delta W_{ij} = \alpha \delta_j Y_i \quad (8)$$

$$W_{jk}(new) = W_{jk}(old) + \Delta W_{ij} \quad (9)$$

$$\Delta V_{ij} = \alpha \delta_j X_i \quad (10)$$

$$V_{jk}(new) = V_{jk}(old) + \Delta V_{ij} \quad (11)$$

D. Algoritma Linear Discriminant Analysis (LDA)

LDA adalah salah satu metode yang dipakai untuk pengenalan pola pada perhitungan statistika dengan cara menemukan proyeksi linear dari data yang akan memaksimalkan jarak antar kelas dan meminimalkan jarak data yang memiliki kesamaan. Metode ini akan menjadi dasar klasifikasi dari data yang ada. Contohnya pengaplikasian dari kombinasi linear ini adalah data buah mangga. Mangga ini dibedakan menjadi mangga manalagi, mangga arumanis, dan mangga kweni. Masing-masing mangga ini dibedakan dari kelasnya dan memiliki nilai jumlah Ni, sebanyak n dimensi (dalam contoh ini karena terdiri dari 3 kelas maka n = 3). Dari data yang ada tersebut, maka sampel dari data tersebut dapat dinotasikan $\{x_1, x_2, \dots, X_{Ni}\}$, merujuk

kepada kelas yang ada. Kemudian data tersebut dikumpulkan menjadi matriks X yang tiap kolom dari matriks ini merepresentasikan dari satu sample data. Setelah itu dicari hasil transformasi dari X ke Y dengan menggunakan proyeksi sample dalam matriks X ke dalam hyperplane dengan dimensi sebesar $K - 1$.¹ (Budiman, Santoso and Afirianto, 2017)

Rumus yang digunakan Dalam melakukan perhitungan LDA ditunjukkan pada Persamaan 3:

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2} \mu_i C^{-1} \mu_i^T + \ln(P_i) \quad (12)$$

- f_i = fungsi diskriminan kelas ke - i
- μ_i = rata-rata nilai setiap kelas dari masing-masing matriks X_i
- C^{-1} = inverse dari grup matriks kovarian
- X_k^T = transpos dari matriks data uji
- P_i = peluang munculnya kelas ke-i
- μ_i^T = transpos rata-rata kelas ke-i
- P_i = peluang munculnya kelas ke-i

E. Confusion Matrix

Confusion Matrix merupakan sebuah hasil evaluasi dari sebuah klasifikasi data mining yang diwujudkan dalam sebuah tabel (Gorunescu, 2011). *Confusion matrix* adalah metode yang banyak dipakai untuk menghitung nilai akurasi. Pengukuran kinerja menggunakan *confusion matrix* memiliki empat istilah sebagai gambaran dari hasil klasifikasi. Adapun keempat istilah tersebut yaitu :

1. *False Positive* (FP), yaitu data negatif tapi terprediksi sebagai data positif.
2. *False Negative* (FN), yaitu data positif yang terprediksi sebagai data negatif.
3. *True Positive* (TP), yaitu data positif yang terprediksi benar.
4. *True Negative* (TN), yaitu data negatif yang terprediksi dengan benar.

dengan klasifikasi yang sebenarnya. Bentuk *Confusion Matrix* secara umum dapat dicermati pada tabel dibawah ini:

Tabel 1. Tabel Confusion Matrix (Gorunescu, 2011)

Classification	Predicted class
----------------	-----------------

		Class : Yes	Class : No
Observed Class	Class Yes	A(True Positive)	B(False Negative)
	Class No	C(False Positive)	D(True Negative)

Untuk menghitung akurasi digunakan rumus sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (13)$$

Algoritma klasifikasi pastinya berusaha untuk menghasilkan model yang menghasilkan akurasi yang baik. Kinerja model dari algoritma klasifikasi ditentukan pada saat model dihadapkan pada data *testing*, karena rata-rata model yang dipakai dapat memprediksi dengan benar pada semua data yang menjadi data *trainingnya*, (Istiawan and Khikmah, 2019).

Sensitivitas atau Recall adalah rasio prediksi benar positif dipadukan dengan keseluruhan data yang benar positif atau mengukur proporsi positif asli yang diramal secara benar sebagai positif. Dalam sensitivitas berkaitan dengan kecakapan pengujian untuk mengenali hasil yang positif dari sejumlah data yang seharusnya positif. Untuk menghitung sensitivitas atau recall menggunakan persamaan dibawah ini:

$$Sensitivitas = \frac{TP}{TP+FN} \quad (14)$$

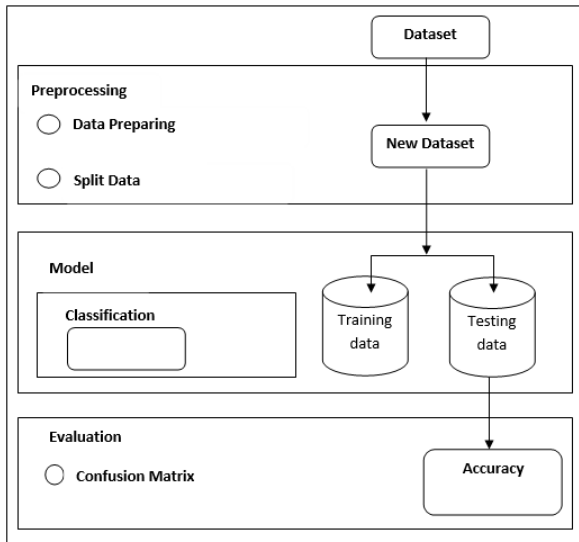
Sedangkan precision adalah rasio ramalan benar positif dipadukan dengan semua hasil yang diprediksi positif. *Precision* menggambarkan matrik untuk menghitung kemampuan sistem dalam menghasilkan data yang penting. *Precision* pada data mining adalah hasil jumlah data yang true positive dibagi dengan jumlah data yang dikenali sebagai positif. Untuk menghitung precision menggunakan persamaan dibawah ini:

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

III. METODE PENELITIAN

Pada penelitian ini akan dilakukan perbandingan algoritma klasifikasi Neural

Network dengan Linier Diskriminant Analysis (LDA) pada klasifikasi penyakit diabetes. Penelitian ini akan menghasilkan manakah hasil akurasi tertinggi, pada saat menggunakan algoritma Neural Network atau LDA. Adapun kerangka penelitian dari penelitian dapat kita cermati pada gambar dibawah ini :



Gambar 2. Kerangka Penelitian

Tahap pertama dalam penelitian ini adalah menyiapkan dataset, selanjutnya di split datanya dibagi menjadi data *testing* dan data *training* dengan presentase 80% untuk data *training* dan 20% untuk data *testing*. Data *training* bertindak sebagai pembentuk pola atau model sedangkan data *testing* bertindak sebagai penguji model. Model yang digunakan adalah klasifikasi dengan algoritma Neural Network atau LDA selanjutnya dievaluasi menggunakan confusion matrix dan menghasilkan akurasi.

A. Sumber Data

Pada penelitian ini menggunakan data dari sumber dataset UCI Machine Learning Repository yang bisa ditemukan di alamat web <https://archive.ics.uci.edu/ml/machine-learning-databases/00529/>. Dataset yang digunakan adalah *Early stage diabetes risk prediction dataset* dimana file tersebut bernama *diabetes_data_upload.csv*.

Variabel yang dipakai pada penelitian ini adalah sebanyak 17 variabel dengan jumlah data sebanyak 520. Ini termasuk data tentang orang-orang termasuk gejala yang dapat menyebabkan diabetes. Kumpulan data ini dibuat dari kuesioner langsung kepada orang-orang yang baru saja menjadi penderita

diabetes, atau yang masih nondiabetes tetapi memiliki sedikit atau lebih gejala. Data dikumpulkan dari pasien dengan menggunakan kuesioner langsung dari Sylhet Diabetes Hospital of Sylhet, Bangladesh.

B. Split Data Otomatis

Pada tahap ini, 520 data diabetes dibagi menjadi dua, yaitu data *training* dan data *testing* dengan prosentase 80% untuk data *training* dan 20% untuk data *testing*. Data *training* bertindak sebagai pembentuk pola atau model dan data *testing* sebagai penguji model.

C. Metode yang Diusulkan

Metode klasifikasi yang diusulkan adalah Neural Network dan LDA.

1. Algoritma Neural Network

Adapun langkah-langkah cara kerja algoritma Backpropagation Neural Network adalah sebagai berikut:

- Langkah pertama cara kerjanya adalah dimulai dengan *input layer*, hitung output dari setiap elemen pemroses melalui *input layer*.
- Hitung error pada *output layer* yang merupakan selisih antara data aktual dan target.
- Transformasikan error tersebut pada bagian yang sesuai di sisi input elemen pemroses.
- Propagasi balik error ini pada output setiap elemen pemroses ke error yang terdapat pada input. Ulangi proses ini sampai input tercapai.
- Ubah seluruh bobot dengan menggunakan kesalahan pada sisi elemen input dan elemen output pemroses yang terhubung.

2. Algoritma LDA

Langkah-langkah algoritma Linear Discriminant Analysis (LDA) yaitu sebagai berikut:

- Membuat matriks rata-rata masing-masing kelas yang diperoleh dari banyaknya kelas yang sama yang kemudian akan digunakan untuk mencari nilai mean corrected yang akan dilakukan untuk langkah selanjutnya.
- Membuat nilai mean global yang diperoleh dari hasil rata-rata keseluruhan dataset.

3. Membuat matriks mean corrected yang diperoleh dari hasil pengurangan data training di masing masing kelas dengan nilai rata-rata mean global.
4. Membuat matriks transpose yang diperoleh dari mean corrected dan kemudian dirubah menjadi matriks transpose.
5. Membuat matriks kovarian perkelas, kovarian grup dan invers. Matriks kovariann diperoleh dari perkalian antara matriks transpose dengan matriks mean corrected dan dibagi oleh jumlah data training di tiap-tiap kelas. Kemudian hasil dari matriks kovarian dijadikan matriks kovarian grup yang diperoleh dari perhitungan hasil penjumlahan kovarian dibagi jumlah data yang digunakan. Tahap selanjutnya mencari matriks invers dari kovarian grup.
6. Membuat Probabilitas Prior di setiap kelas Nilai probabilitas prior diperoleh dari membagi jumlah data training di satu kelas dengan jumlah data keseluruhan.
7. Mencari Fungsi Discriminant Proses fungsi discriminant digunakan untuk mencari nilai tertinggi pada masing-masing kelas yang diujikan dan sebagai hasil akhir dari proses Linear Discriminant Analysis.

IV. HASIL DAN PEMBAHASAN

Variabel data penelitian yang digunakan pada penelitian ini dapat dilihat pada tabel 2 yakni sebagai berikut.

Tabel 2. Variabel data penelitian

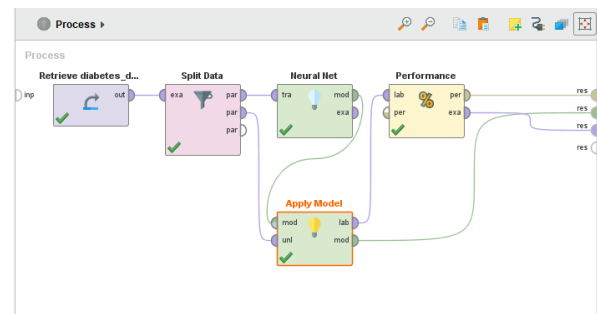
No.	Atribut	Value
1	Umur	1) 20–35, 2) 36–45, 3) 46–55, 4) 56–65, 5) diatas 65
2	Jenis Kelamin	1.Pria, 2.Wanita
3	<i>Polyuria</i>	1.Ya, 2.Tidak
4	<i>Polydipsia</i>	1.Ya, 2.Tidak
5	<i>Suddenweight loss</i>	1.Ya, 2.Tidak
6	<i>Weakness</i>	1.Ya, 2.Tidak
7	<i>Polyphagia</i>	1.Ya, 2.Tidak
8	<i>Genital thrush</i>	1.Ya, 2.Tidak
9	<i>Visual blurring</i>	1.Ya, 2.Tidak
10	<i>Itching</i>	1.Ya, 2.Tidak
11	<i>Irritability</i>	1.Ya, 2.Tidak
12	<i>Delayed healing</i>	1.Ya, 2.Tidak

13	<i>Partial paresis</i>	1.Ya, 2.Tidak
14	<i>Muscle stiffness</i>	1.Ya, 2.Tidak
15	<i>Alopecia</i>	1.Ya, 2.Tidak
16	Obesitas	1.Ya, 2.Tidak
17	Kelas	1.Positif 2.Negatif

Dari Tabel 2 Ada 16 variabel dataset gejala dan 1 variabel class penentu klasifikasi. Selanjutnya dari jumlah data 520, data dibagi menjadi dua dengan presentase 80% data *training* dan 20% data *testing*. 80% dari jumlah data 520 didapatkan data *training* dengan jumlah 416 dan 20% dari jumlah data 520 didapatkan data *testing* dengan jumlah 104. Dari data yang diperoleh selanjutnya dilakukan tahap klasifikasi dengan Neural Network dan LDA.

- 1) Klasifikasi dengan algoritma Neural Network

Penerapan dan pengujian dataset dilakukan pada rapidminer. Berikut ini merupakan proses klasifikasi penderita biabetes dengan klasifikasi algoritma Neural Network menggunakan *RapidMiner 9.7*:



Gambar 3. Proses Klasifikasi Neural Network Menggunakan Aplikasi *Rapidminer 9.7*.

Tahap pertama dalam penelitian ini adalah menyiapkan dataset, selanjutnya di spilt datanya dibagi menjadi data *testing* dan data *training* dengan presentase 80% untuk data *training* dan 20% untuk data *testing*. Selanjutnya dilakukan klasifikasi menggunakan algoritma *Neural Network* dan dihitung akurasi menggunakan *Confusion Matrix*. Dari hasil pengujian menghasilkan *performance vector* dan tabel *Confusion Matrix* seperti dibawah ini:

Tabel 3. Performance Vector Klasifikasi Neural Network

PerformanceVector

```

PerformanceVector:
accuracy: 95.19%
ConfusionMatrix:
True:  Positive      Negative
Positive:    61       2
Negative:    3       38
precision: 92.68% (positive class: Negative)
ConfusionMatrix:
True:  Positive      Negative
Positive:    61       2
Negative:    3       38
recall: 95.00% (positive class: Negative)
ConfusionMatrix:
True:  Positive      Negative
Positive:    61       2
Negative:    3       38
AUC (optimistic): 0.969 (positive class: Negative)
AUC: 0.969 (positive class: Negative)
AUC (pessimistic): 0.969 (positive class: Negative)
    
```

Tabel 4. Hasil Akurasi Klasifikasi Neural Network

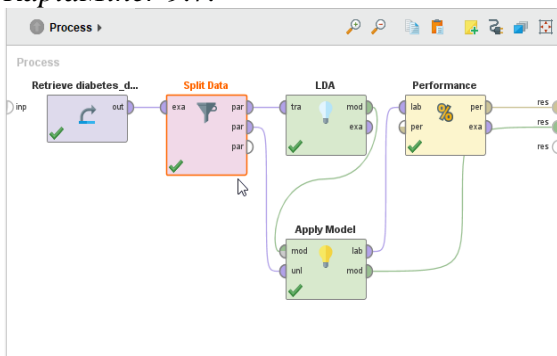
Accuracy: 95.19%

	True Positive	True Negative	Class Precision
Pred. Positive	61	2	96.83 %
Pred. Negative	3	38	92.68 %
Class recall	95.31%	95.00%	

Pengujian menghasilkan akurasi yang cukup besar yaitu 95,19 % Precision sebesar 92,68% %, dan Recall sebesar 95,00%.

2) Klasifikasi dengan algoritma LDA

Penerapan dan pengujian dataset dilakukan pada rapidminer. Berikut ini merupakan proses klasifikasi penderita biabetes dengan klasifikasi algoritma LDA menggunakan *RapidMiner 9.7*:



Gambar 4. Proses Klasifikasi LDA Menggunakan Aplikasi *Rapidminer 9.7*.

Tahap pertama dalam penelitian ini adalah menyiapkan dataset, selanjutnya di spilt datanya dibagi menjadi data *testing* dan data *training* dengan presentase 80% untuk data *training* dan 20% untuk data *testing*. Selanjutnya dilakukan klasifikasi menggunakan algoritma LDA dan dihitung akurasinya menggunakan *Confusion Matrix*. Dari hasil pengujian menghasilkan

performance vector dan tabel *Confusion Matrix* seperti dibawah ini:

Tabel 5. Performance Vector Klasifikasi LDA

PerformanceVector

```

PerformanceVector:
accuracy: 90.38%
ConfusionMatrix:
True:  Positive      Negative
Positive:    56       2
Negative:    8       38
precision: 82.61% (positive class: Negative)
ConfusionMatrix:
True:  Positive      Negative
Positive:    56       2
Negative:    8       38
recall: 95.00% (positive class: Negative)
ConfusionMatrix:
True:  Positive      Negative
Positive:    56       2
Negative:    8       38
AUC (optimistic): 1.000 (positive class: Negative)
AUC: 0.000 (positive class: Negative)
AUC (pessimistic): 0.000 (positive class: Negative)
    
```

Tabel 6. Hasil Akurasi Klasifikasi LDA

Accuracy: 90.38%

	True Positive	True Negative	Class Precision
Pred. Positive	56	2	96.55 %
Pred. Negative	8	38	82.61 %
Class recall	87.50%	95.00%	

Dari tabel diatas bisa dilihat bahwa pengujian menghasilkan akurasi yang cukup besar yaitu 90.38 % Precision sebesar 82,61% %, dan Recall sebesar 95,00%

V. KESIMPULAN

Penelitian pertama menggunakan algoritma klasifikasi Neural Network untuk prediksi seseorang terkena penyakit diabetes atau tidak. Dari 520 data dibagi menjadi 416 sebagai data *training* dan 104 sebagai data *testing*. Dari hasil Pengujian menghasilkan akurasi yang cukup besar yaitu 95,19 % Precision sebesar 92,68% %, dan Recall sebesar 95,00%.

Penelitian kedua menggunakan algoritma klasifikasi LDA untuk prediksi seseorang terkena penyakit diabetes atau tidak. Dari 520 data dibagi menjadi 416 sebagai data *training* dan 104 sebagai data *testing*. Dari hasil Pengujian menghasilkan akurasi yang cukup besar yaitu 90.38 % Precision sebesar 82.61 % %, dan Recall sebesar 95,00%.

Dari hasil penelitian membuktikan bahwa menggunakan algoritma Neural Network untuk klasifikasi penyakit diabetes menghasilkan akurasi lebih besar yaitu

sebesar 95,19 % daripada saat menggunakan LDA.

DAFTAR PUSTAKA

- ADA (2016) 'Classification and Diagnosis of Diabetes Mellitus', *Standards of Medical Care in Diabetes*, 39(January). doi: 10.1016/B978-0-323-18907-1.00038-X.
- Apostolakis, J. (2010) *An introduction to data mining, Structure and Bonding*. doi: 10.1007/430_2009_1.
- Budiman, E., Santoso, E. and Afirianto, T. (2017) 'Pendeteksi Jenis Autis pada Anak Usia Dini Menggunakan Metode Linear Discriminant Analysis (LDA)', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(7), pp. 583–592. Available at: Duhita, W.
- Dita Merawati, R. (2019) 'Penerapan Data Mining Penentu Minat Dan Bakat Siswa Smk Dengan Metode C4 . 5', *Jurnal Algor*, 1(1), pp. 28–37.
- Gorunescu, F. (2011) *Data Mining: Concepts, Models, and Techniques*. Springer.
- Hadianto, N., Novitasari, H. B. and Rahmawati, A. (2019) 'Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network', *Jurnal Pilar Nusa Mandiri*, 15(2), pp. 163–170. doi: 10.33480/pilar.v15i2.658.
- Hana, F. M. (2020) 'Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4 . 5'.
- Istiawan, D. and Khikmah, L. (2019) 'Implementation of C4.5 Algorithm for Critical Land Prediction in Agricultural Cultivation Areas in Pemali Jratun Watershed', *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), p. 67. doi: 10.24014/ijaidm.v2i2.7569.
- Khairani (2012) 'Pengetahuan Diabetes Mellitus Dan Upaya Pencegahan Pada Lansia Di Lam Bheu Aceh Besar', *Pengetahuan Diabetes Mellitus Dan Upaya Pencegahan Pada Lansia Di Lam Bheu Aceh Besar*, 3(3), pp. 58–66.
- Larose, D. T. and Larose, C. D. (2014) *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. doi: 10.1002/9781118874059.
- Maulidi, R. et al. (2018) 'Penerapan Neural Network Backpropagation untuk Klasifikasi Artikel Clickbait', *Seminar Nasional FST 2018*, 1(July), pp. 751–757.
- Nugraha, K. A. et al. (2013) 'Algoritma Backpropagation Pada Jaringan Saraf Tiruan Untuk Pengenalan Pola Wayang Kulit', *Seminar Nasional Informatika*, 2013(semnasIF), pp. 8–13. Available at: Wayang Kulit, Jaringan Saraf Tiruan, Backpropagation, Deteksi Tepi.
- Nurlina (2019) 'Jurnal Media Keperawatan : Politeknik Kesehatan Makassar Jurnal Media Keperawatan : Politeknik Kesehatan Makassar', *Jurnal Media Keperawatan: Politeknik Kesehatan Makassar*, 10(01), pp. 59–66.
- Putra, P. P. and Chan, A. S. (2018) 'Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor)', *INOVTEK Polbeng - Seri Informatika*, 3(1), p. 24. doi: 10.35314/isi.v3i1.296.
- Ramdhani, Y. (2015) 'Komparasi Algoritma LDA Dan Naïve Bayes Dengan Optimasi Fitur Untuk Klasifikasi Citra Tunggal Pap Smear', *Informatika*, II(2), pp. 434–441.
- Rifai, B. (2013) 'Algoritma Neural Network Untuk Prediksi', *Techno Nusa Mandiri*, IX(1), pp. 1–9.
- Suwarno and AA Abdillah (2017) 'Penerapan Algoritma Bayesian Regularization Backpropagation Untuk Memprediksi Penyakit Diabetes', *Jurnal MIPA*, 39(2), pp. 150–158.