

Prediksi Penyakit Diabetes menggunakan Algoritma Artificial Neural Network

Cut Maisyarah K¹, Emy Haryatmi², Riezka Yunistika Fajriatifah³, Yuditha Hatur Puspita⁴

¹Manajemen Sistem Informasi, Fakultas Teknologi Informasi, Universitas Gunadarma

²Teknik Elektro, Fakultas Teknologi Informasi, Universitas Gunadarma

³Manajemen Sistem Informasi, Fakultas Teknologi Informasi, Universitas Gunadarma

⁴Manajemen Sistem Informasi, Fakultas Teknologi Informasi, Universitas Gunadarma

¹csyarah@staff.gunadarma.ac.id, ²emy_h@staff.gunadarma.ac.id, ³riezkeyunistikaf@gmail.com, ⁴yhpuspita@gmail.com

Abstract

Diabetes mellitus is a metabolic disease that occurs when the sugar content in the blood is too high accompanied by metabolic abnormalities as a result of hormonal disorders. High blood sugar levels can cause health problems such as heart disease, nerve damage, reduced vision, and kidney disease. Based on the latest data published in the International Diabetes Federation (IDF) in 2019 shows that 463 million adults currently have diabetes and it is estimated that this number will continue to increase by 51% to 700 million people by 2045. More than 50% of people with diabetes have not been diagnosed, so that it is threatened to develop unnoticed and without prevention. This reveals the urgent need to diagnose diabetics as early as possible in order to provide appropriate and timely treatment. The purpose of this study is to build a model for predicting diabetes using the CRISP-DM method and using an Artificial Neural Network algorithm. This study uses a dataset derived from Pima Indians Diabetes provided by the "National Institute of Diabetes and Digestive and Kidney Diseases" and managed by Kaggle. The data set used will be divided into training data, testing data and validation data, in this study the dataset is divided by a ratio of 70:20:10, 70% for the training process, 20% for testing and 10% validation. In this study, the model that has been built has an 88% accuracy rate for predicting diabetes.

Keywords: Artificial Neural Network, CRISP-DM, data science, diabetes, prediction

Abstrak

Diabetes melitus adalah penyakit metabolik yang terjadi ketika kandungan gula di dalam darah terlalu tinggi yang disertai dengan kelainan metabolik sebagai akibat dari gangguan hormonal. Kadar gula di dalam darah yang tinggi dapat menimbulkan masalah kesehatan seperti sakit jantung, kerusakan saraf, berkurangnya penglihatan, dan penyakit ginjal. Berdasarkan data terakhir yang diterbitkan dalam *International Diabetes Federation* (IDF) tahun 2019 menunjukkan bahwa 463 juta orang dewasa saat ini mengidap diabetes dan diperkirakan angkanya akan terus naik sebanyak 51% menjadi 700 juta orang pada tahun 2045. Lebih dari 50% penderita diabetes belum terdiagnosa, sehingga terancam berkembang tanpa disadari dan tanpa pencegahan. Ini mengungkapkan kebutuhan mendesak untuk mendiagnosis sedini mungkin penderita diabetes agar dapat memberikan perawatan yang tepat dan tepat waktu. Tujuan penelitian ini adalah membangun sebuah model untuk prediksi penyakit diabetes dengan menggunakan metode CRISP-DM dan menggunakan algoritma *Artificial Neural Network*. Penelitian ini menggunakan dataset yang berasal dari Pima Indians Diabetes yang disediakan oleh "*National Institute of Diabetes and Digestive and Kidney Diseases*" dan dikelola oleh Kaggle. Set data yang digunakan akan dibagi menjadi data *training*, data *testing* dan data validasi, dalam penelitian ini dataset dibagi dengan rasio 70:20:10, 70% untuk proses *training*, 20% untuk *testing* dan 10% validasi. Pada penelitian ini model yang sudah dibangun memiliki tingkat akurasi sebesar 88% untuk memprediksi penyakit diabetes.

Kata kunci: *Artificial Neural Network, CRISP-DM, data science, diabetes, prediksi.*

1. Pendahuluan

Diabetes melitus adalah penyakit metabolik yang terjadi ketika kandungan gula didalam darah terlalu tinggi yang disertai dengan kelainan metabolik sebagai akibat dari gangguan hormonal. Kadar gula di dalam darah yang tinggi dapat menimbulkan masalah kesehatan seperti sakit jantung, kerusakan saraf, berkurangnya penglihatan, dan penyakit ginjal. Berdasarkan data terakhir yang diterbitkan dalam *International Diabetes Federation* (IDF) tahun 2019 menunjukkan bahwa 463 juta orang dewasa saat ini mengidap diabetes dan diperkirakan angkanya akan terus naik sebanyak 51% menjadi 700 juta orang pada tahun 2045. Dua pertiga dari penderita diabetes tinggal di daerah perkotaan dan tiga dari empat adalah usia kerja. Pada tahun 2019, lebih dari empat juta orang berusia 20-79 tahun diperkirakan meninggal karena mengidap diabetes. Lebih dari satu juta anak-anak dan remaja memiliki diabetes tipe 1 dan diperkirakan penderita diabetes di atas usia 65 tahun berjumlah 136 juta orang. Jumlah ini akan terus meningkat setiap tahunnya.

International Diabetes Federation (IDF) menunjukkan bahwa hiperglikemia pada kehamilan (HIP) mempengaruhi sekitar satu dari enam kehamilan. Lebih dari 50% penderita diabetes belum terdiagnosa, sehingga terancam berkembang tanpa disadari dan tanpa pencegahan. Ini mengungkapkan kebutuhan mendesak untuk mendiagnosis sedini mungkin penderita diabetes agar dapat memberikan perawatan yang tepat dan tepat waktu.

Saat ini sudah terdapat beberapa inovasi dalam bidang teknologi untuk memprediksi sedini mungkin penyakit diabetes berdasarkan riwayat kesehatan pasien pasien sebelumnya. Salah satu cara untuk memprediksi penyakit diabetes adalah dengan menggunakan *Machine Learning*, kumpulan data yang didapat dari hasil diagnosis pasien pasien sebelumnya dapat diolah dengan menggunakan *data mining*.

Supervised learning merupakan salah satu teknik *data mining* yang dapat digunakan untuk melakukan prediksi. Teknik *supervised learning* membutuhkan *training* dataset dengan menentukan nilai input untuk menghasilkan output sebagai target. terdapat banyak algoritma yang termasuk dalam *supervised learning*, dalam penelitian ini akan menggunakan algoritma *Artificial Neural Network* (ANN).

Artificial Neural Network (ANN) merupakan salah satu algoritma yang terinspirasi oleh cara kerja jaringan saraf biologis di otak manusia. ANN mencoba untuk memformulasikan model matematis sel-sel otak manusia yang terdiri dari elemen pemrosesan yang disebut neuron dan jaringan di antara mereka yang disebut dengan koefisien atau bobot. Karakteristik model ini ditentukan oleh

arsitektur jaringan, fungsi aktivasi, dan proses pelatihan. Arsitektur jaringan menempatkan neuron pada lapisan yang terdiri dari *input layer*, *output layer*, dan *hidden layer* jika ada, yang jumlahnya satu atau lebih. Keunggulan dari ANN untuk memprediksi diagnosis medis adalah tidak membutuhkan keilmuan statistika untuk melakukan proses *training*, kemampuan untuk mendeteksi hubungan *non-linier* yang kompleks antara variabel *dependen* dan *independen*, kemampuan untuk mendeteksi interaksi yang mungkin antara variabel prediktor, dan dapat dikembangkan dengan beberapa algoritma untuk proses *training*.

Beberapa penelitian sebelumnya terkait prediksi penyakit diabetes telah dilakukan. Pada tahun 2020, Resky Ramadhani Santoso, Rani Megasari dan Yudi Ahmad Hambali membuat penelitian yang berjudul “Implementasi Metode Machine Learning Menggunakan Algoritma Evolving Artificial Neural Network Pada Kasus Prediksi Diagnosis Diabetes” penelitian ini menggunakan algoritma EANN untuk meningkatkan akurasi dari algoritma ANN. Set data yang digunakan dalam penelitian ini adalah Pima Indian Diabetes dari *UCI Machine Learning Repository*. Hasil akurasi terbaik dari model yang diusulkan sebesar 83,55%. [1]

Pada tahun 2021, Eky Cahya Putra Witjaksana, Rd. Rohmat Saedudin dan Vandha Pradwiyasma Widartha, membuat penelitian yang berjudul “Perbandingan Akurasi Algoritma Random Forest dan Algoritma Artificial Neural Network untuk klasifikasi penyakit diabetes” penelitian ini membandingkan hasil akurasi dari algoritma random forest dan algoritma ANN dengan membuat *Confusion Matrix* untuk menemukan hasil dari ROC AUC dan hasil F1-Score dari Algoritma yang digunakan. Hasil akurasi yang didapat dari model yang diusulkan menunjukkan algoritma *random forest* mendapatkan nilai akurasi yang lebih baik sebesar 90,62% dibanding dengan algoritma ANN yang memiliki nilai akurasi sebesar 82,29%. [2]

Pada tahun 2022, Hovi Sohibil Wafa, Asep Id Hadiana, dan Fajri Rakhmat Umbara, membuat penelitian berjudul “Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)” penelitian ini menggunakan algoritma SVM *Radial Basis Function* (RBF). Pengujian akurasi pada penelitian ini dilakukan menggunakan *Confusion Matrix* dan peramalan *Mean Square Error* dengan kfold kelipatan 10. penelitian ini mencapai hasil akurasi sebesar 91%. [3]

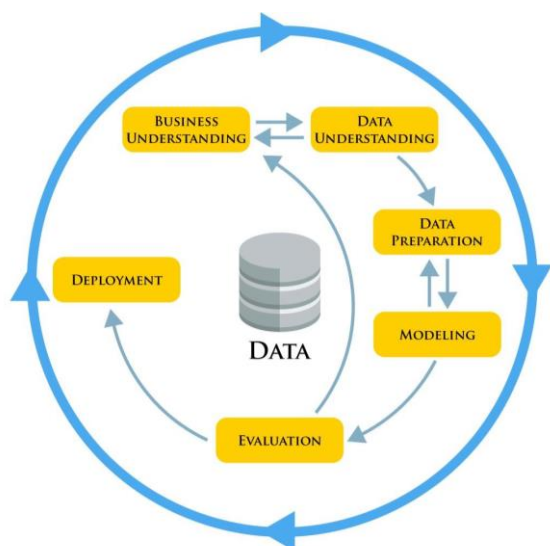
Dataset yang digunakan dalam penelitian ini berasal dari Pima Indians Diabetes yang disediakan oleh “*National Institute of Diabetes and Digestive and Kidney Diseases*” dan dikelola oleh *Kaggle*. Penelitian ini menggunakan dataset ini karena dataset

ini memiliki kredibilitas yang tinggi sehingga layak untuk digunakan sebagai bahan penelitian.

Berdasarkan seluruh paparan diatas, mulai dari penjelasan penyakit diabetes, algoritma yang akan digunakan dan beberapa tinjauan literatur dari penelitian sebelumnya, maka dalam penelitian ini akan mengimplementasikan algoritma *Artificial Neural Network* dengan menggunakan *library sklearn* dalam memprediksi penderita penyakit diabetes dengan harapan model yang diusulkan dapat menghasilkan akurasi yang baik.

2. Metode Penelitian

Dalam melakukan penelitian ini ada beberapa tahapan yang akan dilaksanakan menggunakan metode CRISP-DM atau *cross industry standard process for data mining*. Menurut CRISP-DM, *data mining* memiliki enam fase yang bersifat adaptif, yaitu kegiatan antar fase sangat bergantung satu sama lain. Ketergantungan paling signifikan ditunjukkan oleh panah. Keenam fase tersebut yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* dan *Deployment*.



Gambar 1. Metode CRISP-DM

1. Business Understanding (Pemahaman Bisnis)

Dalam tahapan ini ada beberapa hal yang dilakukan seperti memahami kebutuhan serta tujuan dari bisnis yang selanjutnya akan didefinisikan ke dalam data mining, kemudian menentukan rencana dan strategi untuk mencapai tujuan data mining. Adapun yang menjadi masalah dari penelitian ini adalah penyakit diabetes yang menjadi *silent killer*.

2. Data Understanding (Pemahaman Data)

Setelah melalui tahapan pemahaman bisnis, beberapa hal yang harus dilakukan dalam tahapan ini diantaranya mengenali serta memahami data yang

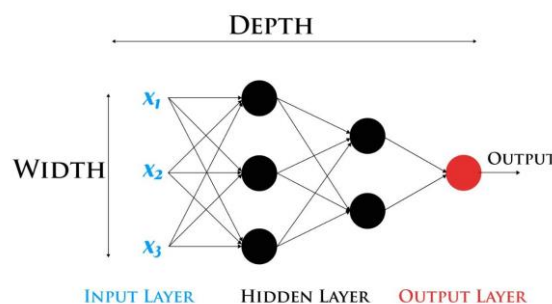
akan digunakan, mengambil atau mengumpulkan data, menelaah data, memvalidasi kualitas dari data.

3. Data Preparation (Persiapan Data)

Pada tahapan ini yang perlu dilakukan adalah memperbaiki kualitas data untuk pemodelan seperti memilih dan memilah data yang akan digunakan, membersihkan data, mengkonstruksi data, dan integrasi data.

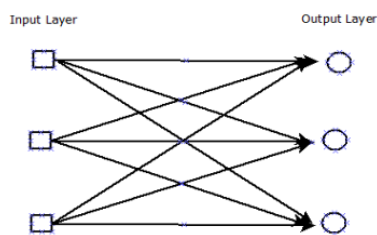
4. Modeling (Pemodelan)

Tahapan ini melibatkan *machine learning* secara langsung untuk penentuan teknik *data mining* yang akan digunakan. Pada tahapan ini membangun skenario pemodelan terbaik dengan memilih algoritma yang akan digunakan, membagi data sesuai dengan ketersediaan data menjadi data latih, data validasi dan data uji. Kemudian membangun model menggunakan algoritma *Artificial Neural Network* (ANN). ANN memiliki banyak hidden layer dan sifat non linier sehingga menghasilkan model yang sulit dibaca dan dimengerti oleh manusia. Sulit membuktikan kebenaran hasil konfigurasi parameter dari proses training. Interpretability model yang rendah menyebabkan sulit untuk menganalisa proses inferensi yang terjadi pada model ANN. Secara matematis, ANN diibaratkan sebagai sebuah graf, ANN memiliki neuron/node dan sinapsis/edge. Karena memiliki notasi seperti graf, notasi aljabar linier lebih mudah digunakan untuk operasi ANN.



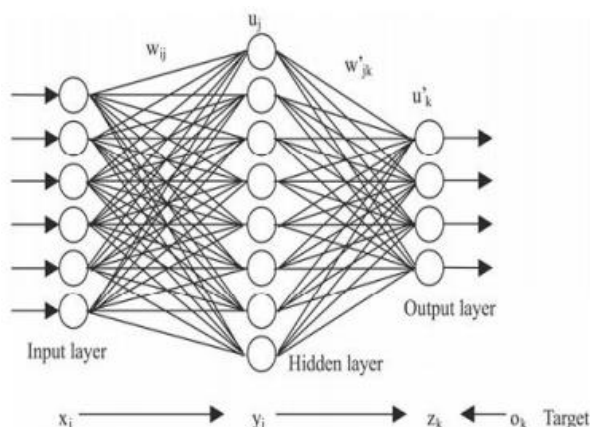
Gambar 2. Arsitektur Ann

Berdasarkan gambar 2 diatas, *depth* pada ANN mengacu pada banyaknya layer, *width* mengacu pada banyaknya unit pada layer. Berdasarkan jumlah Layer arsitektur jaringan *Artificial Neural Network* (ANN) dapat diklasifikasikan menjadi dua kelas yang berbeda yaitu, Jaringan layer tunggal (*single layer network*) dan Jaringan layer jamak (*multi layer network*).



Gambar 3. Jaringan Single Layer Network

Gambar 3 merupakan contoh jaringan *single layer* pada ANN, Jaringan layer tunggal (*single layer network*) merupakan semua unit *input* dalam jaringan ini dihubungkan dalam semua unit *output*, meskipun dengan bobot yang berbeda - beda.



Gambar 4. Jaringan Multi Layer Network

Jaringan layer jamak (*multi layer network*) adalah jaringan layer jamak yang merupakan perluasan dari layer tunggal jaringan layer jamak ini memperkenalkan satu atau layer tersembunyi (*hidden layer*) yang mempunyai simpul yang disebut neuron tersembunyi (*hidden layer*). Contoh dari jaringan *multi layer network* dapat dilihat pada gambar 4.

5. Evaluation (Pengujian)

Tahapan ini dilakukan untuk melihat tingkat performa dari pola yang dihasilkan oleh algoritma yang telah ditentukan. Parameter yang digunakan untuk evaluasi komparasi algoritma adalah *confusion matrix* dengan aturan nilai akurasi, *recall*, presisi dan *F-1 score*. Nilai akurasi menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar. Nilai *recall* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Nilai presisi menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Nilai *F-1 score* menggambarkan perbandingan rata-rata presisi dan *recall* yang dibobotkan. Nilai tersebut dapat diperoleh melalui perhitungan sebagai berikut:

$$\text{Accuracy} = \frac{(\text{Jumlah klasifikasi benar})}{\text{Total sampel testing yang diuji}} \times 100\%$$

$$\text{Recall} = \frac{(\text{True Positive})}{(\text{True Positive} + \text{False Negative})} \times 100\%$$

$$\text{Precision} = \frac{(\text{True Positive})}{\text{True Positive} + \text{False positive}} \times 100\%$$

$$F - 1 \text{ score} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}$$

6. Deployment (Penyebaran)

Tahapan ini merupakan tahapan terakhir yang akan dilakukan, dalam tahap ini akan dilakukan pembuatan laporan dan jurnal ilmiah.

3. Hasil dan Pembahasan

Penelitian ini dilakukan dengan menggunakan metodologi data mining CRISP-DM. metodologi ini menjadi standar dalam pengembangan proyek data mining karena paling banyak digunakan dalam pengembangan data mining. Model CRISP-DM memberikan gambaran siklus hidup proyek data mining dalam 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*.

3.1. Business Understanding

Tujuan bisnis dari penelitian ini adalah untuk merancang model *data mining* yang dapat digunakan sebagai prediktor penyakit diabetes sedari dini karena diabetes menjadi salah satu penyakit yang saat ini menempati posisi ketujuh sebagai penyebab kematian di dunia. Posisi diabetes sebagai *silent killer* harus ditangani dengan baik agar persentase kematian akibat diabetes bisa menurun. Kebanyakan dari penderita diabetes tidak menyadari dirinya menderita diabetes, kebanyakan dari mereka baru mengakses layanan kesehatan ketika kondisinya sudah memburuk sehingga menjadi sulit untuk disembuhkan. hasil dari prediksi ini diharapkan dapat memberikan pencegahan dan pengendalian penyakit diabetes.

3.2. Data Understanding

Penelitian ini menggunakan dataset yang berasal dari Pima Indians Diabetes yang disediakan oleh "National Institute of Diabetes and Digestive and Kidney Diseases" dan dapat diunduh melalui Kaggle. Dataset ini terdiri dari 768 row data dan 9 attribut. Attribut tersebut terdiri dari Pregnancies, Glucose, Blood Pressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, Age, Outcome.

Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1
3	126	88	41	235	39.3	0.704	27	0
8	99	84	0	0	35.4	0.388	50	0
7	196	90	0	0	39.8	0.451	41	1
9	119	80	35	0	29.0	0.263	29	1
11	143	94	33	146	36.6	0.254	51	1
10	125	70	26	115	31.1	0.205	41	1
7	147	76	0	0	39.4	0.257	43	1

Gambar 5. Dataset Diabetes

Dataset ini memiliki 8 variabel independen dan 1 variabel dependen. Tabel 1 akan menjelaskan deskripsi setiap variabel independen beserta tipe data yang digunakan dan tabel 2 akan menjelaskan variabel dependen beserta tipe datanya.

Tabel 1. Variabel Independen

Kolom Dataset	Tipe Data	Deskripsi
Pregnancies	Int64	Variabel ini menunjukkan riwayat kehamilan
Glucose	Float64	Variabel ini menunjukkan hasil tes toleransi glukosa
Blood Pressure	Float64	Variabel ini menunjukkan tekanan darah
Skin Thickness	Float64	Variabel ini menunjukkan perkiraan lemak tubuh yang diukur pada lengan kanan setengah antara proses olecranon dari siku dan proses akromial skapula
Insulin	Float64	Variabel ini menunjukkan tingkat insulin
BMI	Float64	Variabel ini menunjukkan indeks masa tubuh
Diabetes Pedigree Function	Float64	Variabel ini menunjukkan indikator riwayat

		diabetes dalam keluarga
Age	int64	Variabel ini menunjukkan umur

Tabel 2. Variabel Dependen

Kolom Dataset	Tipe Data	Deskripsi
Outcome	Int64	Variabel ini menunjukkan hasil apakah seseorang terkena diabetes atau tidak. Dalam variabel ini sudah dilakukan labeling dengan angka 1 dan 0. 1 menunjukkan hasil terkena diabetes dan 0 tidak terkena diabetes

3.3 Data Preparation

Data preparation terdiri dari *data cleaning*, dan *splitting*. Tahap pertama yang dilakukan dalam penelitian ini adalah *cleaning data* yang bertujuan untuk mencari apakah terdapat data null (*missing value*) ataupun *outlier* didalam dataset. Data *outlier* sangat merugikan karena dapat mempengaruhi proses *learning model* dan dapat mengurangi metrik secara signifikan. Pendeteksian outlier dilakukan satu per satu pada setiap atribut. Setelah dilakukan proses *cleaning data*, jumlah baris dalam dataset yang digunakan berubah dari 768 baris data menjadi 756 baris data.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 756 entries, 0 to 755
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   Pregnancies                          756 non-null    int64
1   Glucose                              756 non-null    float64
2   BloodPressure                        756 non-null    float64
3   SkinThickness                       756 non-null    float64
4   Insulin                             756 non-null    float64
5   BMI                                 756 non-null    float64
6   DiabetesPedigreeFunction             756 non-null    float64
7   Age                                 756 non-null    int64
dtypes: float64(6), int64(2)
memory usage: 47.4 KB
```

Gambar 6. Hasil data cleaning

Tahap selanjutnya merupakan data *splitting*, dataset yang ada akan dibagi menjadi lima bagian, yaitu *x_train*, *x_val*, *x_test*, *y_train* dan *y_test*. *x_train* merupakan variabel yang akan digunakan dalam tahap *training*, *x_test* adalah variabel yang digunakan dalam *testing* dan *x_val* digunakan untuk melakukan validasi. Sedangkan *y_train* merupakan variabel

dependen yang dijadikan sebagai target pada tahap *training*, dan *y_test* akan digunakan dalam tahap *testing*.

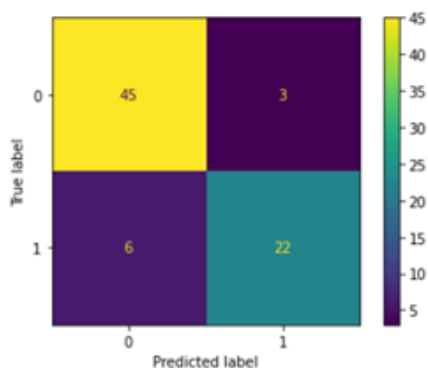
kemudian data akan dibagi menjadi data *training*, data *testing* dan data validasi dalam penelitian ini dataset dibagi dengan rasio 70:20:10, 70% untuk proses *training*, 20% untuk *testing* dan 10% validasi. Proses data *splitting* dilakukan secara acak untuk menghindari terjadinya bias terhadap data *training*, *testing* dan validasi. Setelah proses pembagian data, terdapat 544 baris dan 8 kolom untuk *x_train*, 136 baris dan 8 kolom untuk *x_val*, 76 baris dan 8 kolom untuk *x_test*, 544 baris dan 1 kolom untuk *y_train*, dan 76 baris dan 1 kolom untuk *y_test*.

3.4 Modeling

Teknik pemodelan yang digunakan adalah algoritma *Artificial Neural Network*, implementasi algoritma tersebut menggunakan library scikit-learn dengan bahasa pemrograman Python. Dalam tahapan ini akan dilakukan pemilihan desain uji untuk teknik pemodelan yang akan dipakai, dan dalam penelitian ini metode yang digunakan yaitu *confusion matrix* untuk menghasilkan nilai dari akurasi, presisi, *recall* dan *F1-score*.

3.5 Evaluation (Pengujian)

Dalam tahap evaluasi ini, parameter yang digunakan adalah *confusion matrix* yang pada dasarnya memberikan informasi perbandingan hasil yang telah dilakukan oleh model dengan melihat nilai akurasi, presisi, *recall*, dan *f1-score*. Berikut ini hasil evaluasi pada model yang telah dibuat.

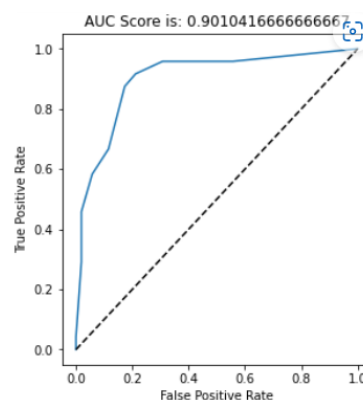


Gambar 7. *Confusion matrix* pada algoritma ANN

Berdasarkan gambar diatas, hasil nilai akurasi adalah accuracy adalah 0.88, presisi = 0.88, recall = 0.79 dan *f1-score* = 0.83. Sehingga didapatkan tingkat akurasi algoritma ANN sebesar 88% dengan jumlah prediksi benar 67 data dari total data testing 76 data.

Setelah mendapatkan hasil perhitungan pada *Confusion matrix* tahapan selanjutnya adalah *ROC curve* berdasarkan antara nilai false positif dengan true positif. *ROC* adalah grafik yang menjadikan hasil

dari false positif sebagai garis horizontal dan hasil dari true positif untuk mengukur perbedaan performansi metode yang digunakan, dan *ROC* biasanya digunakan untuk mengekspresikan *Confusion Matrix*



Gambar 8. Hasil kurva *ROC* algoritma ANN

Dari *ROC curve* diatas, dapat disimpulkan bahwa model yang dibangun memberikan performa yang cukup baik dengan nilai *Area Under the Curve* (AUC) sebesar 0.901 yang berarti baik karena berada diangka 90%.

4. Kesimpulan

Berdasarkan serangkaian proses yang telah dilakukan dalam penelitian ini, maka dapat disimpulkan bahwa penelitian ini berhasil membangun sebuah model untuk prediksi penyakit diabetes menggunakan algoritma *Artificial Neural Network*. Berdasarkan pengujian yang telah dilakukan, proses data cleaning cukup mempengaruhi hasil akurasi dari algoritma *Artificial Neural Network* yang digunakan. Penelitian ini mencapai akurasi terbaiknya sebesar 88%, diharapkan pada penelitian selanjutnya dapat menambah jumlah data sehingga model yang dihasilkan dapat mencapai akurasi yang lebih baik.

Ucapan Terimakasih

Terima kasih kepada *High Performance Computing* (UG-AI-CoE) Universitas Gunadarma dan Program Magister Universitas Gunadarma atas dukungannya terhadap pembuatan jurnal ilmiah ini.

Daftar Rujukan

- [1] Resky, R.S., Rani, M., Yudi, A.H., 2020, Implementasi Metode Machine Learning Menggunakan Algoritma Evolving Artificial Neural Network Pada Kasus Prediksi Diagnosis Diabetes. *Jurnal Aplikasi dan Teori Ilmu Komputer*, 3(2), pp.85-97.
- [2] Eky, C.P.W., Rohmat S., Vandha P.W., 2021. Perbandingan Akurasi Algoritma Random Forest dan Algoritma Artificial Neural Network untuk Klasifikasi Penyakit Diabetes. *e-Proceeding of Engineering*, 8 (5), pp.9773-9781.
- [3] Hovi, S.W., Asep, I.H., Fajri, R.U., 2022. Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). *Informatics and Digital Expert (INDEX)*, 4 (1), pp.40-45.

- [4] A. V. Joshi, 2020, Machine Learning and Artificial Intelligence. Cham, 1st ed, Springer International Publishing.
- [5] Hasdi, P., Nabilah, U.W., 2020. Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. *Jurnal Nasional Teknologi dan Sistem Informasi*, 6 (2), pp.100-107.
- [6] Hilda, A., Kurniati., 2020. Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. *Journal of Information Technology Ampera*, 1(3), pp.133-143
- [7] J. Loy, Neural network projects with Python: the ultimate guide to using Python to explore the true power of neural networks through.
- [8] Msy, A.H., Sopian, S., Ade, S.H., 2021. Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing (JAIC)*, 5 (2), pp.103-108.
- [9] Muhammad, A.W., Windha, M.P., 2022. Optimasi Algoritma Data Mining Menggunakan Backward Elimination Untuk Klasifikasi Penyakit Diabetes. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11 (1), pp.1-12.
- [10] Prana, Y., Raden B.D.P., Eko S.B., 2018. Prediksi Kebangkrutan Menggunakan Artificial Neural Network. *Jurnal Riset Komputer (JURIKOM)*, 5 (5), pp.503-510.
- [11] Supriyono Pangribowo, 2020. “Infodatin Pusat Data Dan Informasi Kesehatan Republik Indonesia”
- [12] Marseba S., Anjar W., Zulaini M.N., 2019. Architectural Model of Backpropagation ANN for Prediction of Population-Based on Sub-Districts in Pematangsiantar City. *International Journal of Information System & Technology*, 3(1), pp.98-106.
- [13] Nuraysah, Z.P., Anjar, W., Ika, O.K., 2019. Implementation of ANN for Prediction of Unemployment Rate Based on Urban Village in 3 Sub-Districts of Pematangsiantar. *International Journal of Information System & Technology*, 3(1), pp.107-116.
- [14] Muhammad, F.M., Urooj A., Muhammad A., Haseeb A., Muhammad Z., 2019. Neural Network Techniques for Time Series Prediction: A Review. *International Journal on Informatics Visualization*, 3(3), pp.314-320.
- [15] Zufian, A., 2021. Artificial Neural Network Model For Wind Mill. *International Journal of Engineering, Science & InformationTechnology (IJESTY)*, 1(3), pp.40-48.