Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

Penerapan Model Decision Tree Menggunakan Python Untuk Prediksi Faktor Dominan Penyebab Penyakit Stroke

Baiq Andriska Candra Permana[1*], Muhammad Sadali[2], Ramli Ahmad[3]
[1,2]Program Studi Teknik Informatika, Universitas Hamzanwadi
[3]Program Studi Teknik Komputer, Universitas Hamzanwadi
*andriska.cp@hamzanwadi.ac.id

Abstrak

Penyakit stroke dikenal sebagai penyakit yang diakibatkan oleh pecahnya pembuluh darah di otak atau penyumbatan pembuluh darah di otak. Gangguan pasokan darah diotak tersebut dapat mengakibatkan rusaknya sel-sel otak yang berakibat rusaknya fungsi otak. Menurut World Health Organization (WHO) stroke merupakan salah satu penyakit penyebab kematian tertinggi diseluruh dunia, sebab hampir 85% penyebab kematian didunia diantaranya adalah karena stroke. Dengan perkembangan ilmu pengetahuan dan teknologi saat ini perlu kiranya dilakukan penelitian untuk melakukan analisa terhadap data penderita stroke sehingga melalui penelitan dapat diketahui faktor–faktor dominan yang mengakibatkan terjadinya stroke. Ini merupakan suatu langkah awal yang dapat dilakukan untuk menyikapi permasalahan tersebut sehingga dapat mengurangi resiko kematian yang diakibatkan oleh penyakit stroke. Dalam penelitian ini analisa dilakukan dengan mengimplementasikan algoritma decision tree menggunakan python karena decission tree memiliki akurasi 91% yang baik dalam melakukan prediksi. Dengan adanya pohon keputusan yang terbentuk dari algoritma ini maka akan mudah melihat gejala mulai yang paling beresiko hingga yang memiliki resiko terendah penyebab penyakit stroke.

Kata kunci : Prediksi, Decission Tree, Pyton, Stroke.

Abstract

Stroke is known as a disease caused by rupture of blood vessels in the brain or blockage of blood vessels in the brain. Disruption of the blood supply to the brain can result in damage to brain cells which results in damage to brain function. According to the World Health Organization (WHO) stroke is one of the highest causes of death worldwide, because almost 85% of the causes of death in the world are due to stroke. Current developments in science and technology require research to analyze stroke patient data so that the dominant factors that lead to stroke can be identified. This is an initial step that can be taken to address these problems to reduce the risk of death caused by stroke. In this study, the analysis was carried out by implementing a decision tree algorithm using python because the decision tree has good accuracy 91% in making predictions. With the decision tree formed from this algorithm, it will be easy to see the symptoms, starting from those who are most at risk to those who have the lowest risk of causing stroke.

Keywords : *Prediction, Decission Tree, Pyton, Stroke*

1. Pendahuluan

Penyakit stroke merupakan salah satu penyakit yang paling banyak menyebabkan kematian diseluruh dunia dan berdasarkan data World Helath Organization (WHO) penderita stroke diseluruh dunia mencapai 15 juta disetiap tahunnya, 5 juta diantaranya meninggal dunia dan yang lainnya mengalami cacat permanen.

Banyak faktor yang dapat menyebabkan terjadinya stroke, selain dari beberapa faktor yang memang tidak dapat dirubah seperti usia maupun jenis kelamin seseorang, stroke juga dipengaruhi

Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

oleh kebiasaan fisik bahkan keadaan sosial ekonomi [1]. Melihat fakta bahaya dari penyakit stroke itu sendiri, maka perlu dilakukan penelitian untuk mengetahui faktor-faktor dominan dari sekian banyak faktor yang menyebabkan penyakit stroke, sehingga kedepannya pencegahan bisa lebih mudah dilakukan sehingga dapat mengurangi tingkat resiko kematian maupun cacat permanen yang diakibatkan oleh penyakit ini.

Perkembangan ilmu pengetahuan dan teknologi khususnya pada bidang ilmu kesehatan banyak memanfaatkan pemodelan machine learning yang ditujukan untuk mempermudah dalam melakukan klasifikasi maupun prediksi terhadap suatu penyakit. Banyak penelitian sebelumnya yang menggunakan machine learning dengan komparasi beberapa algoritma untuk melakukan prediksi penyakit stroke, akan tetapi belum ada yang menjabarkan gejala apa yang dominan menyebabkan penyakit stroke.

Penelitian ini menggunakan algoritma decision tree yang banyak digunakan dalam machine learning dan dalam pemrosesan suatu data karena decision tree dianggap memiliki kemampuan memberikan akurasi yang baik dalam melakukan prediksi. Sedangkan Bahasa pemrograman python dimanfaatkan untuk melakukan pengolahan data karena kemampuannya dalam mengolah data dalam jumlah yang besar. Dengan implementasi algoritma decesion tree dengan menggunakan python akan membentuk suatu hasil berupa pohon keputusan yang dengan jelas dapat memberi gambaran faktor dominan penyebab stroke, dan dari hasil akurasi model dapat diketahui tingkat akurasi prediksi dari faktor dominan penyebab stroke itu sendiri.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Penelitian terkait menjadi salah satu acuan penulis dalam melakukan penelitian sehingga penulis dapat memeperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan. Dari penelitian ini, penulis mengangkat beberapa penelitian sebagai referensi dalam memperkaya bahan kajian. Berikut merupakan penelitian beberapa jurnal terkait dengan penelitian yang dilakukan penulis, yaitu;

Penelitian tahun 2023 oleh K. A. Susanto dkk dengan judul “Implementasi Bahasa Python Dalam Menganalisis Pengaruh Rokok Terhadap Risiko Pasien Terkena Penyakit Stroke“. Pada penelitian ini diperoleh hasil berupa merokok memiliki kontribusi yang besar dalam meningkatkan resiko terjadinya penyakit stroke. Terdapat kaitan antara merokok dengan variable-variable lain yang menjadi faktor pendukung terjadinya stroke [2].

Penelitian tahun 2022 oleh Y. Azhar, A. K. Firdausy, dan P. J. Amelia, dengan judul

Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

penelitian "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. Pada penelitian yang dilakukan adalah untuk membandingkan beberapa algoritma untuk mendapatkan akurasi tertinggi. Berdasarkan analisa dan pengolahan data yang telah dilakukan maka didapatkan bahwa akurasi terbaik adalah dengan menggunakan KNN dengan tingkat akurasi 98.63% [3].

Penelitian tahun 2022 oleh D. E. Cahyani, dengan judul penelitian "Penerapan Machine Learning Untuk Prediksi Penyakit Stroke". Pada penelitian ini melakukan perbandingan kinerja beberapa algoritma machine learning. Secara umum algoritma machine learning yang digunakan sudah dapat memprediksi dengan baik penyakit stroke dengan akurasi tertinggi berada pada algoritma nave bayes [4].

Penelitian tahun 2021 oleh B. A. Candra Permana dan I. K. Dewi Patwari dalam penelitian dengan judul "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes". Penelitian ini mengambil sample data pasien penderita sebanyak 520 data pasien dan 14 atribut. Hasil penelitian menunjukan bahwa untuk prediksi penyakit diabetes decision tree memiliki akurasi yang lebih baik jika dibandingkan dengan Naïve Bayes [5].

Penelitian tahun 2020 oleh M. Santoni, N. Chamidah, dan N. Matondang pada penelitian dengan judul "Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada software KNIME". Komparasi 3 algoritma data mining yang dilakukan untuk melihat performa 3 algoritma dalam melakukan prediksi penyakit hipertensi. Dari hasil analisa dan ujicoba pengolahan data diperoleh hasil bahwa performa algortima terbaik adalah algoritma artificial neural network dengna tingkat akurasi 94,7% [6].

Penelitian tahun 2019 oleh I. Hadayani, dalam Jurnal Aplikasi Sain,Informasi,Elektronika dan Komputer yang berjudul "Penerapan Algoritma C4.5 Untuk Klasifikasi Penyakit Disk Hernia dan Spondylolisthesis Dalam Kolumna Vertebralis". Pada penelitiannya menggunakan 310 data diperoleh akurasi dalam kelasifikasi penyakit dengan tingkat akurasi yang baik yaitu 89%. Untuk mendapatkan tingkat akurasi yang lebih baik dibutuhkan data dengan jumlah yang lebih besar pada penelitian ini [7]

Penelitian tahun 2019 oleh Medy dan Alviva, dengan judul penelitian "Implementasi Algoritma C4.5 Menggunakan Python Untuk Klasifikasi Kepuasan Konsumen. Data diperoleh dengan pengambilan kuisioner terhadap 100 peserta. Dengan mengimplementasikan algoritma yang diusulkan maka akan diketahui apa yang menyebabkan konsumen merasa puas maupun tidak puas dengan tingkat akurasi algoritma dalam melakukan prediksi sebesar 70% [8]

Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

2.2. Landasan Teori

1. Data Mining

Data mining merupakan suatu metode penambangan data sehingga dari data-data tersebut dapat menghasilkan suatu informasi [9]

Stroke

Penyakit stroke merupakan penyakit yang menyerang otak, dimana terjadi gangguan fungsi pada otak karena adanya masalah pada pembuluh darah yang bertugas mensuplai darah menuju otak [10]. Gangguan bisa terjadi karena terjadinya pecah pembuluh darah atau karena adanya sumbatan pada pembuluh darah [11]. Umumnya orang yang menderita stroke ditandai oleh tidak berfungsinya sebagian atau seluruh anggota tubuh, gangguan kesadaran dan mengalami masalah dalam berbicara [12]

2. Decision Tree

Decission Tree merupakan salah satu algroritma yang umum digunakan dalam machine learning dimana penggunaannya dipergunakan untuk membuat keputusan berupa struktur seperti pohon untuk memodelkan suatu hasil [13]. Decision tree merupakan salah satu metode klasifikasi dimana pada metode ini terdapat suatu node dan cabangnya. Setiap node merenunjuk pada suatu atribut sementara cabang menunjuk pada nilai atribut [14][15].

3. Python

Python merupakan bahasa pemrograman yang populer untuk digunakan karena sintax nya mudah di ingat dan dipahami [8]. Python merupakan open source yang umum digunakan untuk melakukan pembuatan website maupun pengolahan data dengan jumlah yang besar. Dalam analisis data python dapat melakukan kalkulasi statistik, menganalisa data maupun visualisasi data. Sedangkan pada machine learning python dapat digunakan untuk membuat algoritma untuk modul. Pytbhon memiliki libraries yang umum dan populer digunakan :

a. Transferflow: dibangun oleh google dan biasanya digunakan untuk menulis algritma machine learning.
b. Sckit-learning: merupakan salah satu library terbaik yang dapat bekerja pada data yang kompleks. Library ini terkait dengan NumPy dan Scipy.
c. NumPy: merupakan library python yang di khususkan untuk ilmu komputer atau data matematika.
d. Keras: implementasi dari neural network yang memiliki kemampuan yang baik dalam pemodelan, evaluasi dataset, visualisasi dan masih banyak lagi

3. Metode Penelitian

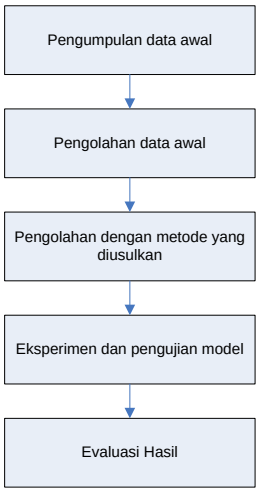
3.1. Tahap Penelitian

Beberapa tahapan yang dilakukan pada penelitian ini yaitu :

Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232



Gambar 1. Tahapan Penelitian

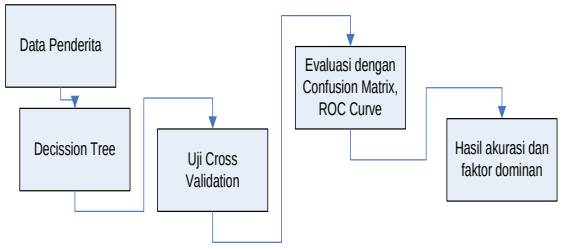
3.2. Pengumpulan Data Awal

Pengumpulan data merupakan tahapan yang dilakukan untuk memperoleh data baik yang didapatkan secara primer yaitu data yang didapatkan langsung oleh peneliti melalui pengumpulan data langsung. Selain itu terdapat juga jenis data skunder dimana data diperoleh dari tangan kedua. Dalam penelitian ini peneliti akan menggunakan data skunder.

3.3. Pengolahan Data Awal

Pengolahan data awal ditujukan untuk menghapus data-data yang tidak konsisten, redundan maupun tidak lengkap.

3.4. Model Yang Diusulkan



Gambar 2. Model Yang Diusulkan

4. Hasil dan Pembahasan

4.1. Data Olahan

Pada Penelitian ini database yang digunakan merupakan data 5110 pasien dengan 12 variable, dimana 11 variable merupakan kondisi klinis pasien dan variable ke 12 merupakan kelas target yang hanya memliliki dua nilai yaitu nilai 1 msnyatakan stroke dan 0 menyatakan tidak stroke.

4.2. Pembahasan dan Implementasi

Pertama import library yang diperlukan dan tabel data:

```
import pandas as pd
import numpy as np
from google.colab import data_table
data_table.enable_dataframe_formatter()
```

Selanjutnya menampilkan tabel data yang di olah

```
dstroke = pd.read_csv("healthcare-dataset-stroke-data.csv")
dstroke
```

index	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
5	56669	Male	81.0	0	0	Yes	Private	Urban	186.21	29.0	formerly smoked	1
6	53882	Male	74.0	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
7	10434	Female	69.0	0	0	No	Private	Urban	94.39	22.8	never smoked	1
8	27419	Female	59.0	0	0	Yes	Private	Rural	76.15	NaN	Unknown	1
9	60491	Female	78.0	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1

Show 10 per page 1 2 10 100 500 510 511

Gambar 3. Cuplikan dataset yang digunakan

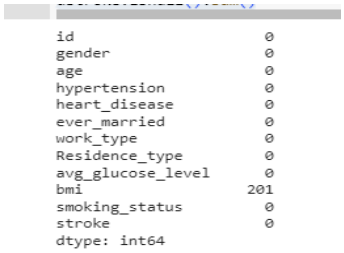
Selanjutnya dari seluruh data dapat dicek data yang imbalance :

Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

```
dstroke = pd.read_csv("healthcare-dataset-stroke-data.csv")
dstroke.isnull().sum()
```



```
id                   0
gender               0
age                  0
hypertension         0
heart_disease        0
ever_married         0
work_type            0
Residence_type       0
avg_glucose_level    0
bmi                201
smoking_status       0
stroke               0
dtype: int64
```

Gambar 4. Pengecekan data yang tidak relevan

Pada gambar terlihat terdapat 201 data yang kurang sesuai atau imbalance. Untuk mempermudah pengolahan data, sebaiknya data dibuat menjadi numerik, karena pada data sebelumnya banyak data yang tersaji dalam string. Beberapa data yang perlu penyesuaian diantaranya : gender, ever_married, work_type, dan smoking status.

Berikut adalah tahapan yang dilakukan pada variable work_type yang terdiri atas 4 kategori pekerjaan dimana privat diberi nilai 0, self-employed nilai 1, gov_job nilai 2 dan anak-anak diberi nilai 3:

```
df['work_type'].unique()
def alter(col):
    if col == 'Private':
        return 0
    elif col == 'Self-employed':
        return 1
    elif col == 'Govt_job':
        return 2
    else:
        return 3

df['work_type'] = df['work_type'].apply(alter)
```

Selanjutnya menghilangkan nilai-nilai Nan yang tidak dapat dioleh :

```
df=pd.read_csv('healthcare-dataset-stroke-data.csv')
df.dropna()
```

Selanjutnya mendeklarasikan variable x dan ya untuk selanjutnya membagi data menjadi data training dan data testing. Pada penelitian ini pembagian digunakan 30 : 70

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 1)# 70% training and 30% test
```

Untuk mendapatkan nilai akurasi decision tree perintah yang diberikan :

```
from sklearn import metrics
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

Hasil akurasi dari algoritma decision tree adalah

```
Accuracy: 0.9106327462491846
```

Gambar 5. Nilai akurasi model decision tree

Evaluasi dengan menggunakan konfusion matriks adalah :

```
from sklearn.metrics import classification_report, confusion_matrix
```

Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

```
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

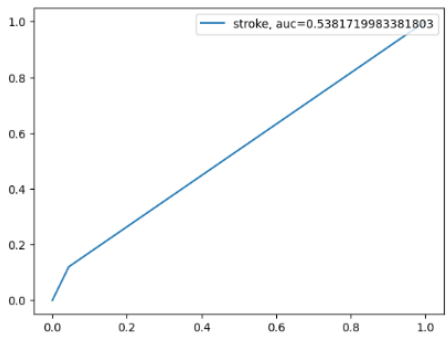
```
[[1386   64]
 [  73   10]]
              precision    recall  f1-score   support

           0       0.95      0.96      0.95      1450
           1       0.14      0.12      0.13        83

    accuracy                           0.91      1533
   macro avg       0.54      0.54      0.54      1533
weighted avg       0.91      0.91      0.91      1533
```

Gambar 6. Hasil Klasifikasi dengan Confusion Matriks

Pada hasil terlihat bahwa setelah dilakukan tes sebanyak 1533 kali jumlah kesalahan prediksi yang dilakukan adalah 137 kali dengan tingkat akurasi algoritma sebesar 91%.

Untuk melihat nilai AUC pada kurva adalah sebagai beritkut :

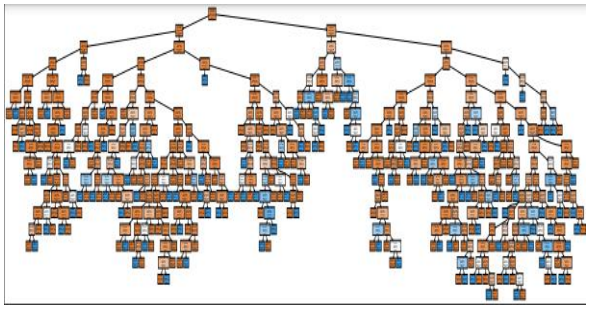
```
import matplotlib.pyplot as plt
y_pred_proba = clf.predict_proba(X_test)[::,1]
fpr, tpr, _ = metrics.roc_curve(y_test,  y_pred)
auc = metrics.roc_auc_score(y_test, y_pred)
plt.plot(fpr,tpr,label="stroke, auc="+str(auc))
plt.legend(loc=1)
plt.show()
```



Gambar 7. Grafik Nilai AUC

Tampilan tree yang dihasilkan pada penelitian ini adalah :

```
dstroke = tree.export_graphviz(clf,
out_file=None,
                feature_names=['gender',
'age', 'hypertension',
'heart_disease','ever_married','work_type','Res
idence_type','avg_glucose_level','bmi','smodin
g_status'],
                class_names='stroke',
                filled=True)
graph = graphviz.Source(dstroke,
format="pdf")
graph.view()
```



Gambar 8. Hasil Tree

Dari tree yang terbentuk dapat ditarik kesimpulan bahwa faktor dominan yang dapat menyebabkan terjadinya stroke dari beberapa variable geja klinis yang muncul adalah bahwa usia, kadar gula darah, ukuran berat badan, hypertensi sebagai faktor utama penyebab terjadinya stroke.

Berdasarkan implementasi yang telah dilakukan terhadap data penyakit stroke yang diolah dengan model data mining menggunakan algoritma c4.5,

Infotek : Jurnal Informatika dan Teknologi
Vol. 7 No. 1, Januari 2024
Hal.23-31

e-ISSN 2614-8773

DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

diketahui bahwa banyak faktor yang mungkin bagi sesorang untuk bisa terkena penyakit stroke. Namun berdasarkan implementasi di temukan bahwa ada faktor-faktor dominan yang menyebabkan stroke tersebut terjadi terutama sekali ketika sesorang memiliki usia dewasa atau lanjut dan tidak menjaga kestabilan kadar gula dalam darahnya.

Tujuan utama dari penelitian ini adalah untuk mengetahui faktor dominan penyebab penyakit stroke, sehingga seseorang bisa mawas diri untuk mengindari terjadinya stroke sejak dini terutama jika mengetahui ada masalah kadar gula dalam darahnya

5. Kesimpulan

Dari hasil dan pembahasan penelitian ditarik kesimpulan bahwa model decision tree memiliki akurasi yang baik dalam memprediksi faktor dominan penyebab stroke yaitu dengan tingkat akurasi 91% dan nilai AUC 0.5. Dari penelitian ini di dapatkan bahwa faktor yang paling berpengaruh sebagai penyebab stroke yang utama adalah usia yang dipicu oleh beberapa faktor lain seperti kadar gula, berat badan dan hypertensi

6. Daftar Pustaka

[1] V. Azzahra and S. Ronoatmodjo, “Faktor-faktor yang Berhubungan dengan Kejadian Stroke pada Penduduk Usia ≥15 Tahun di Provinsi Daerah Istimewa Yogyakarta (Analisis Data Riskesdas 2018),” *J. Epidemiol. Kesehat. Indones.*, vol. 6, no. 2, 2023, doi: 10.7454/epidkes.v6i2.6508.

[2] K. A. Susanto *et al.*, “Implementasi Bahasa Python Dalam Menganalisis Pengaruh Rokok Terhadap Risiko Pasien Terkena Penyakit Stroke,” vol. 2, no. 2, 2023.

[3] Y. Azhar, A. K. Firdausy, and P. J. Amelia, “Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke,” *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 2, pp. 191–197, 2022, doi: 10.31598/sintechjournal.v5i2.1222.

[4] D. E. Cahyani, “Penerapan Machine Learning Untuk Prediksi Penyakit Stroke,” *J. Kaji. Mat. dan Apl.*, vol. 3, no. 1, p. 15, 2022, doi: 10.17977/um055v3i12022p15-22.

[5] B. A. Candra Permana and I. K. Dewi Patwari, “Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes,” *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.

[6] M. M. Santoni, N. Chamidah, and N. Matondang, “Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada software KNIME,” *Techno.Com*, vol. 19, no. 4, pp. 353–363, 2020, doi: 10.33633/tc.v19i4.3872.

[7] I. Handayani, “Penerapan Algoritma C4.5 Untuk Klasifikasi Penyakit Disk Hernia Dan Spondylolisthesis Dalam Kolumna Vertebralis,” *JASIEK (Jurnal Apl. Sains, Informasi, Elektron. dan Komputer)*, vol. 1, no. 2, pp. 83–88, 2019, doi: 10.26905/jasiek.v1i2.3185.

[8] Medy and Alviva, “Implementasi Algoritma C4.5 Menggunakan Python Untuk Klasifikasi Kepuasan Konsumen,” *Progres*, pp. 49–55, 2019, [Online]. Available: https://jurnal.stmikprofesional.ac.id/index.php/Progress/article/view/146/22.

[9] I. Verawati and M. Wishnu, “Penerapan Data Mining Untuk Rencana Penambahan Stok Produk Menggunakan Algoritma Apriori Data mining,” *J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 1, pp. 128–133, 2021, [Online]. Available: https://doi.org/10.30743/infotekjar.v6i1.3884.

[10] R. M. Bantuan and D. A. N. Tatalaksana, “Pengetahuan tentang stroke, faktor risiko, tanda peringatan, respon mencari bantuan, dan tatalaksana pada pasien stroke iskemik di kota semarang,” vol. 2, no. November, pp. 12–21,

e-ISSN 2614-8773
DOI : 10.29408/jit.v7i1.23232 Link : https://dx.doi.org/10.29408/jit.v7i1.23232

2019.
[11] P. A. W. Suwaryo, W. T. Widodo, and E. Setianingsih, "The Risk Factors That Influence the Incidence of Stroke," *LPPM Sekol. Tinggi Ilmu Kesehat. Kendal*, vol. 11, no. 4, pp. 251–260, 2019.
[12] Y. A. Utama and S. S. Nainggolan, "Faktor Resiko yang Mempengaruhi Kejadian Stroke: Sebuah Tinjauan Sistematis," *J. Ilm. Univ. Batanghari Jambi*, vol. 22, no. 1, p. 549, 2022, doi: 10.33087/jiubj.v22i1.1950.
[13] Noviandi, "Implementasi Algoritma Decision Tree C4.5 Untuk Prediksi Penyakit Diabetes," *Inohim*, vol. 6, no. 1, pp. 1–5, 2018.
[14] B. A. C. Permana, R. Ahmad, H. Bahtiar, A. Sudianto, and I. Gunawan, "Classification of diabetes disease using decision tree algorithm (C4.5)," *J. Phys. Conf. Ser.*, vol. 1869, no. 1, 2021, doi: 10.1088/1742-6596/1869/1/012082.
[15] A. Andriani, "Sistem prediksi penyakit diabetes berbasis decision tree," *J. Bianglala Inform.*, vol. I, no. 1, pp. 1–10, 2013.