Jurnal Komputer Multidisipliner

Vol.9 No.1, Januari 2026
ISSN: 24559633

EFEKTIVITAS ALGORITMA *MACHINE LEARNING* DALAM KLASIFIKASI PENYAKIT HIPERTENSI DENGAN MENGGUNAKAN EDA (STUDI KASUS DINAS KESEHATAN MUSI RAWAS)

Nabila[1], Rudi Kurniawan[2], Joni Karman[3]
Universitas Bina Insan
E-mail: nabilaamin1119@gmail.com[1], rudikurniawan@univbinainsan.ac.id[2], joni_karman@univbinainsan.ac.id[3]
***Coressponding Author: Joni Karman**
joni_karman@univbinainsan.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis efektivitas algoritma machine learning dalam klasifikasi penyakit hipertensi di Puskesmas Mangunharjo, Kabupaten Musi Rawas. Data yang digunakan mencakup 1.733 data pasien dari periode Januari hingga Desember 2023, di mana 550 di antaranya terdiagnosis hipertensi. Metode pengumpulan data meliputi data primer dari registrasi pasien dan data sekunder yang relevan. Proses analisis dimulai dengan tahap preprocessing untuk menangani missing values, standarisasi, dan normalisasi data. Penelitian ini menerapkan beberapa algoritma klasifikasi, termasuk Random Forest, Logistic Regression, SVM, Naïve Bayes, dan ANN. Evaluasi dilakukan menggunakan teknik Cross-Validation k-fold 10, Confusion Matrix, dan grafik ROC-AUC. Hasil menunjukkan bahwa algoritma ANN mencapai akurasi tertinggi sebesar 95%, diikuti oleh LR dan NB dengan akurasi 91%. Klasifikasi ini berhasil mengidentifikasi 1.998 pasien hipertensi dan 283 pasien non-hipertensi. Dari analisis yang telah dilakukan, diketahui bahwa penerapan machine learning dan teknik EDA efektif dalam meningkatkan akurasi diagnosis hipertensi. Penelitian ini memberikan wawasan penting bagi pengembangan sistem pendukung keputusan di bidang kesehatan, serta menekankan perlunya penerapan teknologi dalam pengelolaan kesehatan masyarakat.

Kata Kunci — *Machine Learning*, Klasifikasi, Hipertensi, Puskesmas Mangunharjo, *Exploratory Data Analysis* (EDA).

Abstract

This research aims to analyze the effectiveness of machine learning algorithms in classifying hypertension at the Mangunharjo Community Health Center, Musi Rawas Regency. The data used includes 1,733 patient records from January to December 2023, of which 550 were diagnosed with hypertension. Data collection methods include primary data from patient registration and relevant secondary data. The analysis process begins with the preprocessing stage to handle missing values, standardization, and normalization of the data. This study applies several classification algorithms, including Random Forest, Logistic Regression, SVM, Naïve Bayes, and AN). Evaluation was conducted using the Cross-Validation k-fold 10 technique, Confusion Matrix, and ROC-AUC graphs. The results show that the ANN algorithm achieved the highest accuracy of 95%, followed by Logistic Regression and Naïve Bayes with 91% accuracy. This classification successfully identified 1,998 hypertensive patients and 283 non- hypertensive patients. From the analysis, it can be concluded that the application of machine learning and EDA techniques is effective in increasing the accuracy of hypertension diagnosis. This research provides important insights for the development of decision support systems in the health sector and emphasizes the need to apply technology in public health management.

Keywords *— Machine Learning, Classification, Hypertension, Mangunharjo Community Health Center, Exploratory Data Analysis (EDA).*

1. PENDAHULUAN

Hipertensi atau tekanan darah tinggi adalah salah satu masalah kesehatan global yang dapat memicu berbagai komplikasi serius seperti penyakit jantung dan stroke. Mengingat jumlah kasus hipertensi yang semakin meningkat, metode yang efektif untuk mengklasifikasikan dan memprediksi penyakit ini sangat diperlukan. Pendekatan yang menjanjikan adalah penggunaan algoritma machine learning.(N. tri Putri et al., 2022)

Data kesehatan nasional menunjukkan bahwa jumlah penderita hipertensi terus meningkat setiap tahunnya, terutama di wilayah perkotaan dan pedesaan yang mengalami perubahan gaya hidup dan pola makan.(Dr. Frits Reinier Wantian Suling Sp.JP(K), FIHA, 2021)

Puskesmas Mangunharjo, sebagai salah satu fasilitas kesehatan tingkat pertama di Kabupaten Musi Rawas , yang menghadapi tantangan serupa. Berdasarkan laporan tahunan, jumlah kunjungan pasien dengan diagnosis hipertensi mengalami peningkatan yang signifikan dalam beberapa tahun terakhir. Hal ini dipengaruhi oleh berbagai faktor, seperti pola makan tidak sehat, kurangnya aktivitas fisik, dan rendahnya kesadaran masyarakat akan pentingnya deteksi dini dan pengelolaan tekanan darah

Namun, penanganan kasus hipertensi di Puskesmas Mangunharjo masih dilakukan secara manual, baik dalam pencatatan data pasien maupun dalam pemantauan perkembangan tekanan darah. Sistem manual ini menimbulkan beberapa kendala, seperti risiko kehilangan data, lambatnya proses pelaporan, dan sulitnya melakukan analisis tren kasus untuk pengambilan keputusan yang berbasis data. Selain itu, edukasi dan intervensi kepada masyarakat juga belum optimal karena terbatasnya sumber daya dan teknologi pendukung.

Pembelajaran machine learning digunakan dalam analisis data kesehatan, dimana algoritma banyak digunakan memprediksi risiko kesehatan dan mendeteksi penyakit sejak dini. Penelitian saat ini menunjukkan bahwa algoritma pembelajaran mesin dapat memberikan kinerja prediktif yang lebih baik daripada model probabilistik dalam berbagai aplikasi.(Purnama Sari et al., 2024)

Algoritma machine learning juga sudah banyak diimplementasikan atau digunakan untuk mendukung deteksi dini dan klasifikasi penyakit secara efektif. Selain itu, Exploratory Data Analysis (EDA) merupakan langkah penting dalam memahami pola data dan meningkatkan kualitas data sebelum proses klasifikasi dilakukan.

2. TINJAUAN PUSTAKA

1. Hipertensi

Hipertensi adalah penyakit yang kompleks yang disebabkan oleh banyak faktor.(Control et al., 2024) Hipertensi merupakan penyakit yang dapat terjadi pada siapa saja, baik pria maupun wanita, dan mungkin tidak memiliki gejala atau tanda mereka yang terkena mungkin tidak menyadari gejalanya dan menyebabkan kerusakan serta komplikasi pada sistem dan organ kardiovaskular tubuh.(Utari et al., 2021)

2. Data Mining

Data mining merupakan penemuan informasi yang tersembunyi dalam database dan merupakan bagian dari (KDD) untuk menemukan informasi dan pola yang berguna dalam data. Data mining melibatkan keterlibatan komputer dan manusia untuk mencari informasi baru, berharga, dan berguna dalam kumpulan data dan dilakukan secara berulang melalui proses otomatis atau manual.(Zahra et al., 2024)

3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah proses menganalisis dan menampilkan data bertujuan mendapatkan pemahaman yang lebih baik tentang wawasan dari

data.(Radhi et al., 2022)

Tujuan EDA adalah mencari pola data. Hal ini sejalan dengan konsep data mining yang juga digunakan untuk mengekplorasi pola dari suatu data. EDA digunakan dalam tujuan pengurangan dimensi data atau memperkaya pemahaman atas analisis data melalui visualisasi data Ada berbagai langkah yang dilakukan saat melakukan EDA.(Wahyuni et al., 2019)

4. Pyhton

Python merupakan bahasa pemrograman dinamis dengan manajemen memori otomatis. Seperti bahasa pemrograman dinamis lainnya, Python umumnya digunakan sebagai bahasa scripting, namun kenyataannya penggunaan bahasa tersebut lebih luas, mencakup konteks aplikasi yang biasanya tidak dimiliki oleh bahasa scripting. Python juga dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi.

5. Confusion Matrix

Confusion Matrix merupakan metode pengukuran untuk mencari masalah klasifikasi yang dilakukan oleh machine learning dengan keluaran berupa dua kelas atau lebih, confusion Matrix berupa tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual.(M. Putri, 2024)

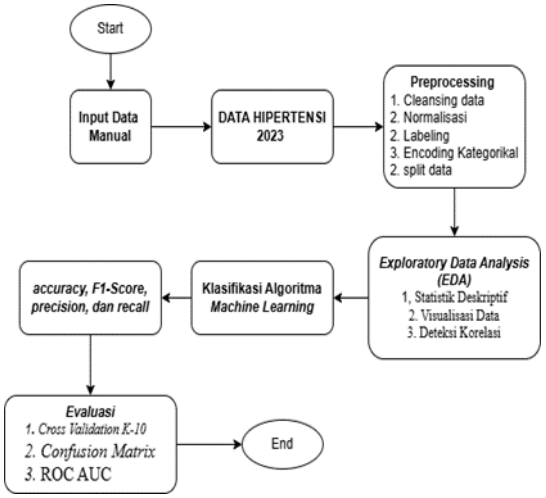
6. K-Fold Cross Validation

K-Fold Cross Validation (CV) adalah teknik statistik yang digunakan untuk mengevaluasi kinerja model prediksi dalam machine learning.(Hafid, 2023)

K-Fold Cross Validation (CV) merupakan cara yang efektif dalam menggabungkan atribut dan setting parameter machine learning dalam melatih model prediksi yang lebih baik.(Wijiyanto et al., 2024).

3. METODELOGI PENELITIAN

Pada penelitian ini menerapkan metode kuantitatif dengan fokus pada analisis data yang dikumpulkan dari Puskesmas Mangunharjo Kabupaten Musi Rawas. Data ini akan diolah melalui klasifikasi menggunakan algoritma machine learning. Alur penelitian dapat dilihat pada gambar ini.



Gambar 1 Kerangka kerja penelitian

1. Metode Pengumpulan Data

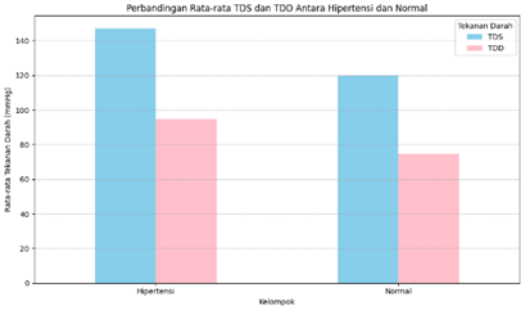
Studi Pustaka

Studi Pustaka adalah proses pengumpulan data dari sumber-sumber literatur. Hal ini dilakukan dengan mengumpulkan materi melalui jurnal, buku, e-book dan situs internet

yang relevan dengan penelitian, khususnya mengenai klasifikasi penyakit hipetrensi menggunakan algoritma machine learning.

Data Primer

Data primer adalah data yang diperoleh secara langsung dari objek penelitian. Dalam penelitian ini, data diambil dari Puskesmas Mangunharjo Kabupaten Musi Rawas, selama periode satu tahun (2023) dengan total sebanyak 1.733 data. Dan data yang dikumpulkan mencakup informasi seperti Nama, Umur, Berat Badan (BB), Tinggi Badan (TB), Tekanan Darah Sistolik (TDS), Lingkar Pinggang (LP), dan Tekanan Darah Diastolik (TDD), Keterangan Diagnosis.



Gambar 2 Grafik perbandingan antara hipertensi dan normal

2. Tempat dan Waktu Penelitian

Tempat Penelitian ini dilakukan di Puskesmas O Mangunharjo dengan data yang diambil adalah data penderita penyakit diabetes. Penelitian ini dilakan pada tahun 2024.

3. Metode Analisa

Penelitian ini menggunakan metode analisis dengan menerapkan algoritma Machine Learning untuk mengevaluasi performa acuracy, precision, recall, serta mengukur ROC AUC. Dengan tujuan utama yaitu mengidentifikasi algoritma yang memberikan kinerja terbaik dalam mengklasifikasikan penyakit hipertensi pada Puskesmas Mangunharjo.

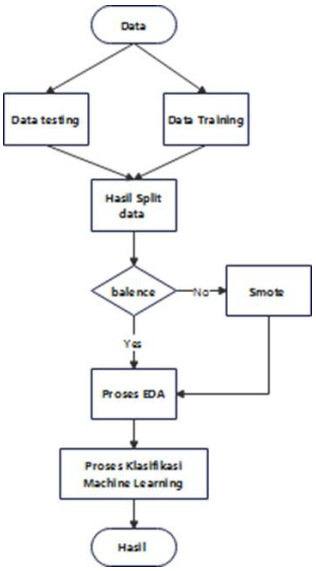
4. Metode Pengajuan Dan Pengolahan Data

Metode Pengujian

Metode pengujian menggunakan Cross Validation K- fold 10, Confusion Matrix, dan ROC AUC sebagai evaluasi algoritma Machine Learning untuk memperesentasikan hasil dari proses klasifikasi.

Pengolahan Data

Pengolahan data dalam penelitian ini melibatkan serangkaian langkah mulai dari preprocessing data hingga penerapan algoritma Machine Learning.



Gambar 3 Tahapan mengolah data

1. Dataset / Data

Jumlah data yang akan digunakan pada penelitian ini sebanyak 1.733 data registrasi pasien dan 550 data real penderita hipertensi . Dan data yang dikumpulkan mencakup informasi seperti Nama, Umur, Berat Badan (BB), Tinggi Badan (TB), Tekanan Darah Sistolik (TDS), Lingkar Pinggang (LP), dan Tekanan Darah Diastolik (TDD), Keterangan Diagnosis penderita.

2. Preprocessing
   a. Split data adalah proses membagi dataset menjadi beberapa bagian untuk keperluan tertentu dalam analisis data atau pengembangan model machine learning.(Muraina, 2022) Data biasanya dibagi menjadi dua bagian, yaitu Data Training dan Data Testing. Perbandingan training dan testing perlu disesuaikan dengan ukuran dataset dan kebutuhan penelitian. seperti 70:30, 80:20, dan 90:10.(Geron, 2017) Dan jika hasil split data seimbang atau balance maka dapat dilanjut ketahapan berikut nya. Dan jika data tidak seimbang atau imbalance maka akan menggunakan metode SMOTE.
   b. SMOTE adalah teknik yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset. Ketidakseimbangan kelas terjadi ketika jumlah contoh dari satu kelas jauh lebih sedikit dibandingkan dengan kelas lainnya, yang dapat menyebabkan model machine learning cenderung bias terhadap kelas mayoritas.(Elreedy et al., 2024)
3. Exploratory Data Analysis (EDA)
   a. Statistik Deskriptif: Dilakukan perhitungan dasar seperti rata-rata, median, standar deviasi, dan lainnya untuk memahami distribusi data.
   b. Visualisasi Data: Menggunakan grafik seperti histogram, scatter plot, atau heatmap untuk memahami pola, outlier, dan hubungan antar variabel.
   c. Deteksi Korelasi: Mengidentifikasi hubungan antara variabel-variabel dalam data. Misalnya, apakah ada korelasi kuat antara tekanan darah dengan usia atau berat badan.
4. Evaluasi
   a. Cross Validation K-10: Data dibagi menjadi 10 subset (fold), dan model dilatih serta diuji sebanyak 10 kali, dengan setiap subset bergantian menjadi data uji. Teknik ini membantu memastikan model tidak overfitting.
   b. Confusion Matrix: Matriks yang menggambarkan performa model dengan

menghitung prediksi benar (true positives, true negatives) dan salah (false positives, false negatives).

c. ROC AUC: ROC (Receiver Operating Characteristic) mengukur kemampuan model dalam membedakan antara dua kelas, sedangkan AUC (Area Under Curve) adalah nilai numerik yang mencerminkan kinerja keseluruhan.

4. HASIL DAN PEMBAHASAN

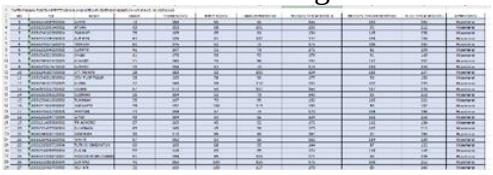
1. Pengumpulan Data

Dengan total data yang dikumpulkan sebanyak 1.733 data pasien. Dan data Real penderita hipertensi atau data pasien yang sudah di diagnosis penyakit hipertensi yang berjumlah 550 data. Jenis data yang dikumpulkan yaitu : Nama Pasien: Identitas pasien, Umur, Berat Badan (BB), Tinggi Badan (TB), Tekanan Darah Sistolik (TDS), Tekanan Darah Diastolik (TDD), Lingkar Pinggang (LP), Gula Darah Sewaktu (GDS), Keterangan Diagnosis: Status

DATA REGISTRASI PASIEN WILAYAH KERJA PUSKESMAS MANGUNHARJO KAB. MUSI RAWAS

NO	NAMA	UMUR	TB	BB	LP	TDS	TDD	GDS
1	SITI MASYENI	73	152	65	91	164	87	338
2	ANANG SULTONI	63	153	88	101	99	117	234
3	RIA NINGSIH	35	159	86	65	114	60	173
4	SAJI	64	165	81	54	133	76	95
5	SRI YANTI	61	145	52	72	128	87	256
6	BAMBANG PUGU	41	167	71	91	139	107	164
7	MUJI LESTARI	41	165	91	100	100	112	260
8	SUMINI	23	153	97	77	132	74	373

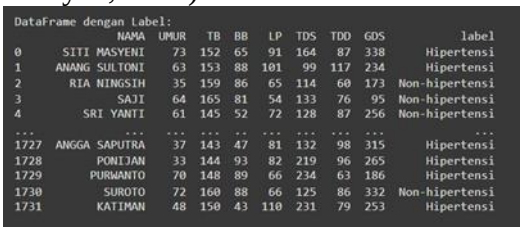
Gambar 4 Data registrasi



Gambar 5 Data Real Hipertensi

2. Labeling

Labeling data adalah proses menambahkan label atau kategori ke setiap data dalam dataset. Tujuan dari labeling data adalah untuk memberikan identifikasi atau penandaan yang jelas terhadap setiap data, sehingga data tersebut dapat digunakan untuk melatih model pembelajaran mesin yang memiliki tujuan tertentu, seperti klasifikasi atau prediksi.(Graciela & Hafiz Irsyad, 2024)



DataFrame dengan Label:

	NAMA	UMUR	TB	BB	LP	TDS	TDD	GDS	label
0	SITI MASYENI	73	152	65	91	164	87	338	Hipertensi
1	ANANG SULTONI	63	153	88	101	99	117	234	Hipertensi
2	RIA NINGSIH	35	159	86	65	114	60	173	Non-hipertensi
3	SAJI	64	165	81	54	133	76	95	Non-hipertensi
4	SRI YANTI	61	145	52	72	128	87	256	Non-hipertensi
...	...	...	...	..	...	...	...	...	...
1727	ANGGA SAPUTRA	37	143	47	81	132	98	315	Hipertensi
1728	PONIJAN	33	144	93	82	219	96	265	Hipertensi
1729	PURWANTO	70	148	89	66	234	63	186	Hipertensi
1730	SUROTO	72	160	88	66	125	86	332	Non-hipertensi
1731	KATIMAN	48	150	43	110	231	79	253	Hipertensi

Gambar 6 Hasil labeling

3. Preprocessing

Tahapan pre-processing dalam analisis data pada data mining dapat dibagi menjadi beberapa sub bagian(Aditya et al., 2024), yaitu :

a) Cleansing Data

Sebuah data yang hilang atau data yang tidak tersedia dalam dataset. Missing values ini bisa berdampak pada performa model yang akan dibuat dikarenakan model akan kesusahan dalam menangani data yang hilang tersebut.

```
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   NAMA    1732 non-null   object
 1   UMUR    1732 non-null   int64
 2   TB      1732 non-null   int64
 3   BB      1732 non-null   int64
 4   LP      1732 non-null   int64
 5   TDS     1732 non-null   int64
 6   TDD     1732 non-null   int64
 7   GDS     1732 non-null   int64
 8   label   1732 non-null   int64
dtypes: int64(8), object(1)
memory usage: 121.9+ KB
None
                UMUR           TB           BB           LP           TDS  \
count  1732.000000  1732.000000  1732.000000  1732.000000  1732.000000
mean     49.648383   160.218822    69.929561    83.004042   159.379330
std      15.214296    11.513789    17.591510    19.792671    44.790028
min      18.000000   140.000000    40.000000    50.000000    90.000000
25%      37.000000   151.000000    55.000000    66.000000   122.000000
50%      50.000000   160.000000    70.000000    83.000000   156.000000
75%      63.000000   169.000000    85.000000   100.000000   195.000000
max      80.000000   199.000000   100.000000   119.000000   249.000000

                TDD          GDS        label
count  1732.000000  1732.000000  1732.000000
mean     82.596998   217.206697     0.763857
std      14.484094    87.042988     0.424834
min      60.000000    70.000000     0.000000
25%      71.000000   141.000000     1.000000
50%      82.000000   217.000000     1.000000
75%      93.000000   291.000000     1.000000
max     120.000000   400.000000     1.000000
```

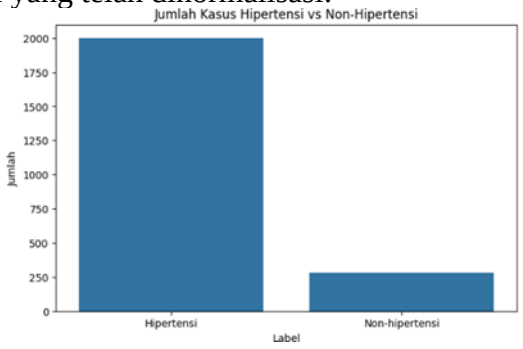
Gambar 7 Hasil Cleansing Data

b) Standarisasi dan Normalisasi

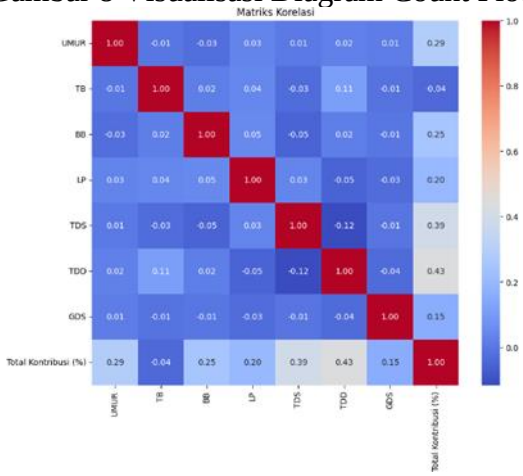
Standarisasi Data Numerik untuk memastikan bahwa semua fitur numerik berada dalam skala yang sama, kolom stok dinormalisasi menggunakan Standard Scaler. Proses ini mengubah nilai-nilai dalam kolom tersebut sehingga memiliki rata-rata 0 dan deviasi standar 1, yang membantu dalam meningkatkan kinerja model.

c) Visualisasi Data

Visualisasi Data berfungsi untuk memahami hubungan antar fitur dalam dataset. Dengan menggunakan matriks korelasi untuk mengidentifikasi kekuatan dan arah hubungan antara variabel yang telah dinormalisasi.



Gambar 8 Visualisasi Diagram Count Plot.



Gambar 9 Visualisasi Matrix korelasi Matriks korelasi yang ditunjukkan dalam heatmap memberikan wawasan penting tentang hubungan antara berbagai fitur dalam dataset, termasuk UMUR, TB, BB, LP, TDS, TDD, GDS, dan Total

Kontribusi (%). Terdapat korelasi positif yang kuat antara TB dan BB, dengan nilai 1.00, yang menunjukkan bahwa peningkatan tinggi badan cenderung diikuti oleh peningkatan berat badan secara proporsional. Di sisi lain, hubungan antara TDS dan TDD menunjukkan korelasi negatif yang lemah (-0.12), menandakan bahwa tidak ada pola yang jelas antara kedua variabel ini.

Selain itu, fitur seperti TDD dan BB menunjukkan tidak adanya hubungan signifikan (0.00), sementara LP dan GDS memiliki korelasi positif yang sangat lemah (0.15). Secara keseluruhan, matriks korelasi ini membantu dalam memahami interaksi antar fitur, yang dapat mendukung proses pemilihan fitur dan pengembangan model machine learning yang lebih efektif, dengan menyoroti bahwa fitur TB dan BB memiliki relevansi yang paling tinggi.

d) Klasifikasi

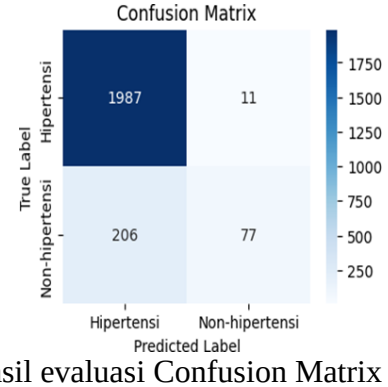
Dalam penelitian ini, klasifikasi dilakukan dengan menerapkan algoritma Support Vector Machine. ANN (Multi Layer Perceptron), Naïve Bayes, dan Logistic Regression, Random Forest. Data dibagi menjadi data pelatihan dan pengujian menggunakan teknik Cross-Validation k- fold 10, di mana satu subset data digunakan sebagai data pengujian, sementara sembilan subset lainnya digunakan sebagai data pelatihan. Proses ini diulang sebanyak 10 kali untuk memastikan evaluasi yang konsisten dan representatif.

e) Evaluasi

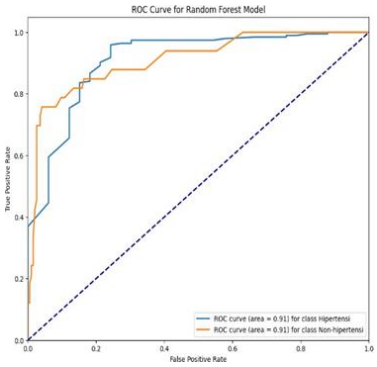
Evaluasi hasil klasifikasi mencakup Cross Validation k-fold 10, Confusion Matrix, dan grafik ROC-AUC. Ini memberikan pemahaman yang mendalam tentang kinerja model yang diterapkan menggunakan Support Vector Machine. ANN (Multi Layer Perceptron), Naïve Bayes, dan Logistic Regression, Random Forest.. Hasil evaluasi dari lima algoritma dapat dilihat sebagai berikut :

```
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.99      0.95      1998
           1       0.88      0.27      0.42       283

    accuracy                           0.90      2281
   macro avg       0.89      0.63      0.68      2281
weighted avg       0.90      0.90      0.88      2281

   Fold  Accuracy  Precision (macro)  Recall (macro)  F1 Score (macro)
0     1  0.934498           0.843889        0.585877          0.625695
1     2  0.899123           0.947727        0.629032          0.677550
2     3  0.925439           0.962222        0.575000          0.610804
3     4  0.881579           0.854893        0.676316          0.720880
4     5  0.894737           0.945701        0.612903          0.655502
5     6  0.877193           0.936652        0.600000          0.632850
6     7  0.921053           0.898864        0.643382          0.697524
7     8  0.929825           0.923611        0.709063          0.770335
8     9  0.907895           0.865312        0.655714          0.705589
9    10  0.877193           0.806033        0.613520          0.647838

Rata-rata: ['0.90', '0.89', '0.63', '0.68']
```

Gambar 10 Hasil evaluasi Cross Validation K-10 Random Forest



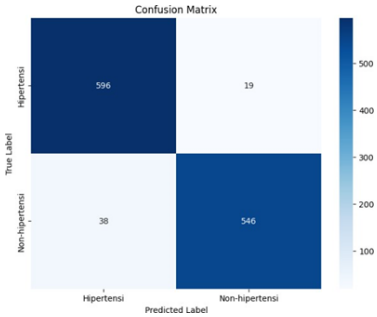
Gambar 11 Hasil evaluasi Confusion Matrix Random Forest



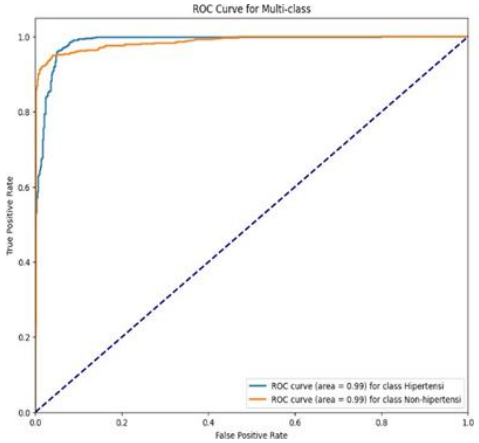
Gambar 12 Grafik ROC Random Forest

```
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.97      0.95       615
           1       0.97      0.93      0.95       584

    accuracy                           0.95      1199
   macro avg       0.95      0.95      0.95      1199
weighted avg       0.95      0.95      0.95      1199

   Fold  Accuracy  Precision (macro)  Recall (macro)  F1 Score (macro)
0     1  0.938865           0.781308        0.750555          0.764818
1     2  0.899123           0.830189        0.683396          0.727202
2     3  0.925439           0.779785        0.687981          0.722330
3     4  0.877193           0.802696        0.705263          0.738661
4     5  0.912281           0.891080        0.704601          0.758218
5     6  0.890351           0.826923        0.701332          0.741555
6     7  0.912281           0.768756        0.748775          0.758218
7     8  0.921053           0.830159        0.737624          0.773610
8     9  0.916667           0.834928        0.737500          0.774645
9    10  0.912281           0.880383        0.734732          0.782940

Rata-rata: ['0.91', '0.82', '0.72', '0.76']
```

Gambar 13 Hasil evaluasi Cross Validation K-10 ANN



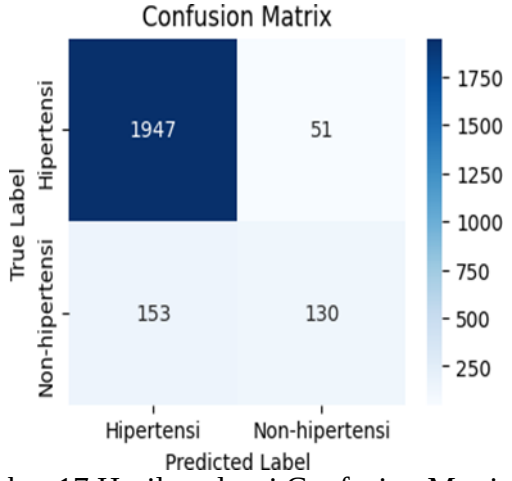
Gambar 14 Hasil evaluasi Confusion Matrix ANN



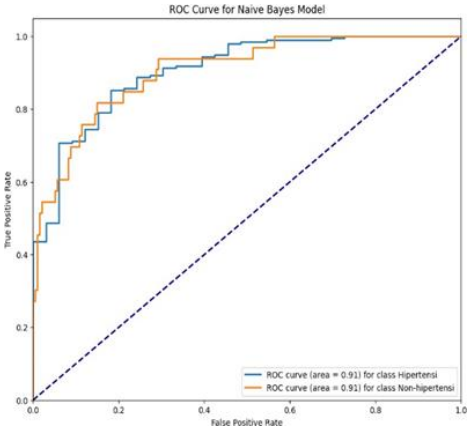
Gambar 15 Grafik ROC ANN

```
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.97      0.95      1998
           1       0.72      0.46      0.56       283

    accuracy                           0.91      2281
   macro avg       0.82      0.72      0.76      2281
weighted avg       0.90      0.91      0.90      2281

   Fold  Accuracy  Precision (macro)  Recall (macro)  F1 Score (macro)
0     1  0.938865           0.781308        0.750555          0.764818
1     2  0.899123           0.830189        0.683396          0.727202
2     3  0.925439           0.779785        0.687981          0.722330
3     4  0.877193           0.802696        0.705263          0.738661
4     5  0.912281           0.891080        0.704601          0.758218
5     6  0.890351           0.826923        0.701332          0.741555
6     7  0.912281           0.768756        0.748775          0.758218
7     8  0.921053           0.830159        0.737624          0.773610
8     9  0.916667           0.834928        0.737500          0.774645
9    10  0.912281           0.880383        0.734732          0.782940

Rata-rata: ['0.91', '0.82', '0.72', '0.76']
```

Gambar 16 Hasil evaluasi Cross Validation K-10 NB



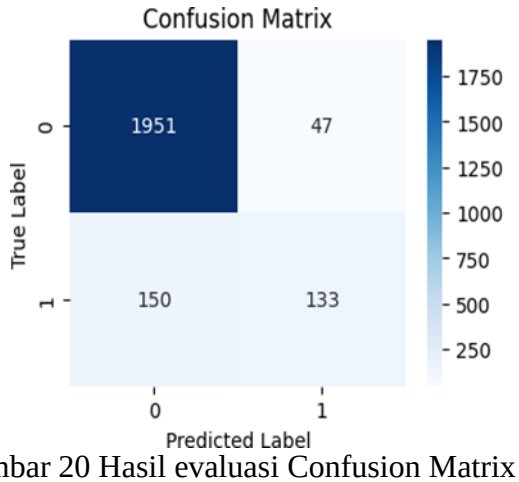
Gambar 17 Hasil evaluasi Confusion Matrix NB



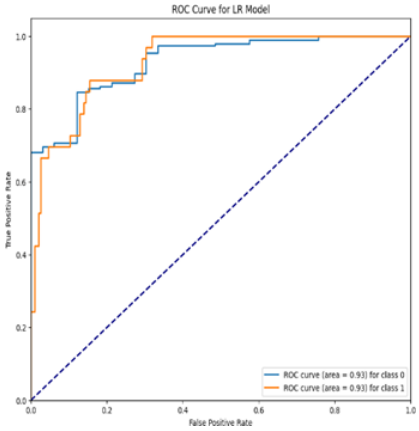
Gambar 18 Grafik ROC NB

```
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.98      0.95      1998
           1       0.74      0.47      0.57       283

    accuracy                           0.91      2281
   macro avg       0.83      0.72      0.76      2281
weighted avg       0.91      0.91      0.91      2281

   Fold  Accuracy  Precision (macro)  Recall (macro)  F1 Score (macro)
0     1  0.951965           0.951965        0.951965          0.951965
1     2  0.899123           0.899123        0.899123          0.899123
2     3  0.921053           0.921053        0.921053          0.921053
3     4  0.872807           0.872807        0.872807          0.872807
4     5  0.921053           0.921053        0.921053          0.921053
5     6  0.881579           0.881579        0.881579          0.881579
6     7  0.925439           0.925439        0.925439          0.925439
7     8  0.929825           0.929825        0.929825          0.929825
8     9  0.907895           0.907895        0.907895          0.907895
9    10  0.925439           0.925439        0.925439          0.925439

Rata-rata: ['0.91', '0.83', '0.72', '0.76']
```

Gambar 19 Hasil evaluasi Cross Validation K-10 LR



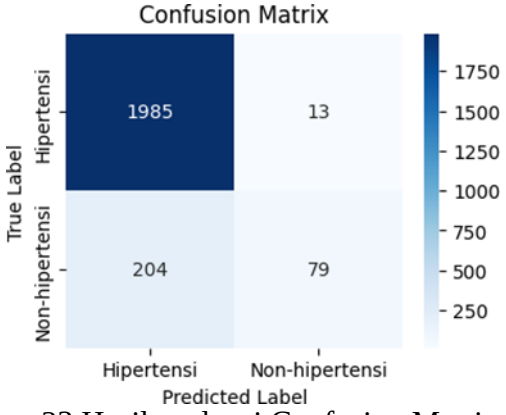
Gambar 20 Hasil evaluasi Confusion Matrix LR



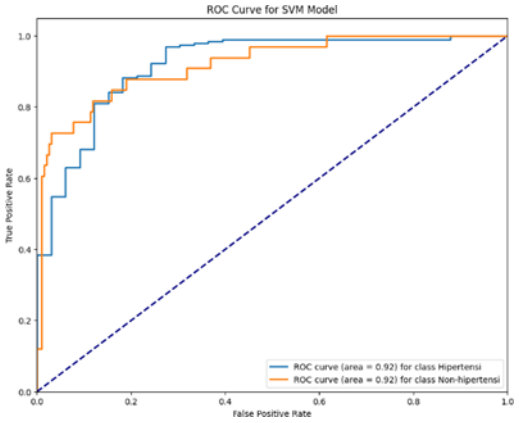
Gambar 21 Grafik ROC LR

```
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.99      0.95      1998
           1       0.86      0.28      0.42       283

    accuracy                           0.90      2281
   macro avg       0.88      0.64      0.68      2281
weighted avg       0.90      0.90      0.88      2281

   Fold  Accuracy  Precision (macro)  Recall (macro)  F1 Score (macro)
0     1  0.943231           0.889761        0.644700          0.702449
1     2  0.894737           0.945701        0.612903          0.655502
2     3  0.934211           0.859209        0.670192          0.723815
3     4  0.877193           0.895833        0.642105          0.685517
4     5  0.894737           0.945701        0.612903          0.655502
5     6  0.877193           0.936652        0.600000          0.632850
6     7  0.921053           0.959459        0.625000          0.678873
7     8  0.907895           0.822162        0.646420          0.691157
8     9  0.899123           0.801968        0.650714          0.691801
9    10  0.899123           0.879506        0.676690          0.727202

Rata-rata: ['0.90', '0.88', '0.64', '0.68']
```

Gambar 22 Hasil evaluasi Cross Validation K-10 SVM



Gambar 23 Hasil evaluasi Confusion Matrix SVM



Gambar 24 Grafik ROC SVM Permasalah yang dibahas dalam

penelitian ini adalah menganalisis klasifikasi penyakit hipertensi dalam kategori Hipertensi dan Non-hipertensi dengan menggunakan Exploratory Data Analysis dan Machine Learning melalui penerapan algoritma Random Forest Classifier, Support Vector Machine, ANN (Multi Layer Perceptron), Naïve Bayes, dan Logistic Regression.

Tujuan penelitian ini adalah untuk membandingkan kinerja algoritma machine learning tersebut dalam klasifikasi. Penelitian ini dimulai dengan mengumpulkan data pasien Puskesmas Mangun Harjo yang mana terdapat 2 sumber yaitu Data Registrasi atau Data

Pasien Pengunjung pada Puskesmas Mangun Harjo selama periode tahun 2023 berdasarkan observasi didapatkan sebanyak 1773 data pasien yang belum di diagnosis penyakit Hipertensi dan Data Real Hipertensi atau Data pasien yang sudah didiagnosis penderita penyakit hipertensi sebanyak 550 data pasien. Setelah data terkumpul langkah selanjut nya adalah melakukan prepocessing, hasil data dapat dilihat pada tabel dibawah ini

Tabel 1 Klasifikasi Random Forest

Kategori	*Precision*	*Recall*	*F1 Score*	*Accuracy*	ROC AUC	CV K-10
Hipertensi	0.91	0.99	0.95	0.90	0.91	0.90
Non hipertensi	0.88	0.27	0.42			

Tabel 2 Klasifikasi Logistic Regression

Kategori	*Precision*	*Recall*	*F1 Score*	*Accuracy*	ROC AUC	CV K-10
Hipertensi	0.93	0.98	0.95	0.91	0.93	0.91
Non hipertensi	0.74	0.47	0.57			

Tabel 3 Klasifikasi Naive Bayes

Kategori	*Precision*	*Recall*	*F1 Score*	*Accuracy*	ROC AUC	CV K-10
Hipertensi	0.93	0.97	0.95	0.91	0.91	0.91
Non hipertensi	0.72	0.46	0.56			

Tabel 4 Klasifikasi ANN

Kategori	*Precision*	*Recall*	*F1 Score*	*Accuracy*	ROC AUC	CV K-10

Hipertensi	0.94	0.97	0.95	0.95	0.99	0.91
Non hipertensi	0.97	0.49	0.95			

Tabel 5 Klasifikasi Support Vector Machine (SVM)

Kategori	*Precision*	*Recall*	*F1 Score*	*Accuracy*	ROC AUC	CV K-10
Hipertensi	0.94	0.97	0.95	0.95	0.99	0.85
Non hipertensi	0.97	0.49	0.95			

5. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa pencatatan data pasien dan pemantauan tekanan darah secara manual di Puskesmas Mangunharjo menimbulkan risiko kehilangan data dan kesulitan analisis tren. Untuk mengatasi masalah ini, penerapan algoritma Machine Learning dan teknik Exploratory Data Analysis (EDA) diperlukan. Dari 2281 data pasien yang dianalisis, 1998 terdiagnosis hipertensi dan 283 non-hipertensi, dengan EDA membantu memahami karakteristik data sebelum penerapan algoritma.

Model klasifikasi yang paling efektif adalah Artificial Neural Network (ANN) dengan akurasi tertinggi 95%, diikuti oleh Logistic Regression dan Naive Bayes yang masing-masing mencapai akurasi 91%. Meskipun Random Forest dan Support Vector Machine memiliki akurasi 90%, keduanya menghadapi tantangan dalam mendeteksi pasien hipertensi. Penelitian ini menegaskan bahwa penerapan teknologi Machine Learning, terutama ANN, dapat meningkatkan akurasi pemantauan kesehatan dan memberikan pemahaman tentang pentingnya teknologi dalam pengelolaan kesehatan yang lebih efektif.

Saran

Agar dapat mengingkatkan hasil penelitian di tahap selanjutnya, ada beberapa rekomendasi yang dapat diajukan kepada peneliti berikutnya. Rekomendasi- rekomendasi ini diharapkan dapat membantu penelitian mencapai hasil yang

lebih optimal, yaitu:

1. Peningkatan Data dan Variabel: Disarankan untuk mengumpulkan lebih banyak data dan mempertimbangkan variabel tambahan yang mungkin berpengaruh terhadap hipertensi, seperti faktor genetik, riwayat kesehatan keluarga, dan kebiasaan hidup lainnya.
2. 2. Penerapan Metode SMOTE: Sebagai langkah untuk menangani ketidakseimbangan data, disarankan untuk menerapkan metode Synthetic Minority Oversampling Technique (SMOTE) dalam proses preprocessing data. Ini dapat meningkatkan akurasi model klasifikasi.
3. Edukasi dan Pelatihan: Mengadakan pelatihan bagi tenaga medis mengenai penggunaan teknologi dan algoritma machine learning dalam klasifikasi penyakit, sehingga meningkatkan pemahaman dan penerapan hasil penelitian di lapangan.
4. Analisis Lanjutan: Mengembangkan analisis lanjutan yang tidak hanya berfokus pada klasifikasi, tetapi juga prediksi risiko jangka panjang bagi pasien hipertensi, sehingga dapat memberikan rekomendasi yang lebih komprehensif.

DAFTAR PUSTAKA

N. tri Putri, R. R, N. Febrianti, andS. S, “Faktor-faktor yang Berhubungan dengan Kejadian Hipertensi pada Ibu Hamil,” An Idea Nurs. J., vol. 1, no. 01, pp. 43–50,2022, doi:

10.53690/inj.v1i01.114.
Fa. Dr. Frits Reinier Wantian Suling Sp.JP(K), FIHA, Fakultas Kedokteran Universitas Kristen Indonesia, vol. 8, no. 2. 2021.
E. Purnama Sari et al.,. “Studi Literatur Deep Learning dan Machine Learning untuk Analisis dan Prediksi Pasar Saham: Metodologi, Representasi Data dan Studi Kasus,” J. Teknol. dan Sains Mod., vol. 1, no. 1, pp. 19–28, 2024, [Online]. Available: https://journal.scitechgrup.com/index.php/jtsm
D. Control et al., “Copyright © 2024 Wolters Kluwer Health , Inc . All rights reserved . Copyright © 2024 Wolters Kluwer Health , Inc . All rights reserved .,” vol. 38, no.. 3, pp. 430–434, 2024.
R. Utari, N.. Sari, and F. E. Sari, “Efektivitas Pendidikan Kesehatan perhadap Motivasi Diit Hipertensi Pada Lansia Hipertensi di Posyandu Lansia Desa Makarti Tulang Bawang Barat Tahun 2020,” J. Dunia Kesmas, vol. 10, no. 1, pp. 136–144,2021, doi: 10.33024/jdk,.v10i1.3550.
R. Adolph,. No Title No Title No Title,” vol. 6, pp. 1–23, 2016.
F. Zahra, M. A. Ridla, and N,. Azise, “Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap),” JUSTIFY J,. Sist. Inf. Ibrahimy, vol. 3, no. 1, pp. 55–65, 2024, doi: 10.35316/justify.v3i1.5335.
M. Radhi, A. Amalia, D. R. H. Sitompul, S. ,H. Sinurat, and E. Indra, “Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook,.”J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA), vol. 4, no. 2, pp. 23–27, 2022, doi: 10.34012/jurnalsisteminformasidanil mukomputer.v4i2.2475.
E. D. Wahyuni, A.. A. Arifiyanti, and M. Kustyani, “Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining,” Pros. Nas. Rekayasa Teknol. Ind. dan Inf. XIV Tahun 2019, vol. 2019, no. November, pp. 263–269, 2019, [Online]. Available: http://journal.itny.ac.id/index.php/R eTII
M. Putri, “Prediksi Penyakit Stroke Menggunakan Machine Learning Dengan Algoritma Random Forest,” J. Infomedia Tek. Inform., vol.. 9, no. 2, pp. 16–21, 2024.
S. Dewi,. “Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan,” Techno Nusa Mandiri, vol. 13, no. 1, pp. 60–66, 2016.
A. Vanacore, M. S. Pellegrino, and A. Ciardiello, “Fair evaluation of classifier predictive performance based on binary confusion matrix,” Comput.. Stat., vol.. 39, no. 1, pp. 363–383, 2024, doi: 10.1007/s00180-022-01301-9.
H. Hafid,/ “Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia,” J. Math., vol. 6, no. 2, pp. 161–168, 2023,. [Online]. Available: http://www.ojs.unm.ac.id/jmathcos
W. Wijiyanto., A. I. Pradana, S. Sopingi, and V. Atina, “Teknik K- Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa,” J. Algoritm., vol. 21, no. 1, pp. 239– 248, 2024, doi: 10.33364/algoritma/v/.21-1.1618.
R. P. Pridiptama, W. Wasono, and F. D., . Amijaya, “Perbandingan Algoritma Support Vector Machine dan Na¨ıveNa¨ıve Bayes pada Klasifikasi Penyakit Tekanan Darah Tinggi (Studi Kasus: Klinik Polresta Samarinda),” Basis., vol. 3, no. 1, pp. 1–16, 2024, [Online]. Available: http://jurnal.fmipa.unmul.ac.id/index .php/Basis
M. F. R. Aditya, N. Lutvi, and U. Indahyanti,/ “Prediksi Penyakit Hipertensi Menggunakan Metode Decison Tree dan Random Forest,” J. Ilm. Komputasi,. vol. 23, no. 1, pp. 9–16, 2024, doi: 10.32409/jikstik.23.1.3503.
K. Jenis,. H. Berdasarkan, F. Pribadi, P. Dewi, P. Purwono, and S. D. Kurniawan, “Pemanfaatan Teknologi Machine Learning pada,” pp. 377–387/.
A. Az., R. Septian,M. A. Saktiawan, and R. A. Saputra, “PREDIKSI PENYAKIT HIPERTENSI MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA REGRESI LOGISTIK,” vol. 8, no. 6, pp., 12062–12068, 2024..
U., Nijunnihayah, S. S. Hilabi, F. Nurapriani, and E. Novalia, “Implementasi Algoritma K-Nearest

Neighbor untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 2, pp. 695–701, 2024,/ doi: 10.57152/malcom.v4i2./1326.

I. O. Muraina, "Ideal Dataset Splitting Ratios in Machine Learning Algorithms:General Concerns for Data Scientists and Data Analysts," 7th Int. Mardin Artuklu Sci. Res. Conf., no. February, pp. 496–504, 2022, [Online].. Available: https://www.researchgate.net/publication/358284895_IDEAL_DATASET_SPLITTING_RATIOS_IN_MACHINE_LEARNING_ALGORITHMS_GENERAL_CONCERNS_FOR_DATA_SCIENTISTS_AND_DATA_ANALYSTS

A. Geron, Hands-On Machine Learning with/ Scikit-Learn & tensorflow: O'Reilly Media, Inc. 2017.

D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," Mach. Learn., vol. 113,. no. 7, pp. 4903–4923, 2024, doi: 10.1007/s10994-022-06296-4.

M. Graciela and dan Hafiz Irsyad, "Klasifikasi Opini Masyarakat Terhadap Naturalisasi Pemain Sepak Bola Menggunakan KNN dan SMOTE," Aicoms, vol. 3, no. 1, pp. 21–27, 2024,. [Online]. Available: https://jurnal.politap.ac.id/index.php/aicoms

K. Aditya, A.. Wisnu, and A. M. A. Rahim, "Analisis Perbandingan Algoritma XGBoost Dan Algoritma Random Forest Untuk Klasifikasi Data Kesehatan Mental," vol. 2, no. 5, pp. 808–818, 2024.