|  |  |
| --- | --- |
|  | **Mathematical Modelling of Engineering Problems**  Vol., No., Month, Year, pp. \*\*-\*\*  Journal homepage: |

|  |  |  |  |
| --- | --- | --- | --- |
| **Identification of Hypertension Using Artificial Neural Network (A Case Study of Modopuro Mojokerto Health Center)** | | |  |
| Muhammad Muqtafin Nuha1, Eka Mala Sari Rochman1,2 | | |  |
| 1Departemen of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia  2 Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Indonesia  Corresponding Author Email: | | | |
| https://doi.org/10.18280/mmep.xxxxxx |  | **ABSTRACT** | |
|  |  |  | |
| **Received:**  **Accepted:** |  | Penelitian ini mengeksplorasi tantangan penuaan yang melibatkan perubahan emosional, seksual, sosial, kognitif, dan fisik, dengan focus pada isu hipertensi. Meskipun hipertensi merupakan masalah serius pada orang lanjut usia, deteksi dini sering sulit karena gejalanya yang tidak selalu terlihat. Dengan menggunakan metode purposive sampling, peneliti memilih 1526 pasien hipertensi dari Puskesmas Modopuro Mojokerto. Data yang dianalisis mencakup berbagai variabel, seperti jenis kelamin, usia, kebiasaan merokok, konsumsi alcohol, pola makan, tinggi, berat badan, indeks massa tubuh (IMT), tekanan darah, laju nafas, detak jantung, dan Diagnosa 1. Penelitian dilakukan dari September hinggan Desember 2023 dengan menerapkan analisis Knowledge Discovery in Database (KDD), yang melibatkan langkah-langkah seperti input data, pre-processing, balancing, sharing, klasifikasi, dan output. Berbagai metode klasifikasi, termasuk Artificial Neural Network, Naïve Bayes dan Decision Tree, dievaluasi dan dibandingkan untuk hasil klasifikasi optimal. Hasilnya menunjukkan bahwa Artificial Neural Network (ANN) memiliki akurasi rata-rata tertinggi sebesar 93,7%, mengungguli Naïve Bayes dan Decision Tree masing-masing dengan 4% dan 2,8%, tanpa penerapan Synthetic Minority Over-sampling Technique (SMOTE). Meskipun demikian, penelitian ini memberikan saran untuk pengembangan lebih lanjut dengan menambah jumlah data, mempertimbangkan factor-faktor tambahan, dan melakukan validasi eksternal untuk memperkuat temuan ini dalam konteks populasi yang lebih luas. | |
| ***Keywords:***  *Hipertensi, Artificial Neural Network (ANN), Naïve Bayes, Decision Tree, SMOTE.* |  |

# Introduction

Peningkatan jumlah orang tua memberikan perhatian khusus terhadap kondisi penuaan mereka seiring dengan permasalahan yang berkaitan dengan proses penuaan berupa perubahan emosional, seksual, social, kognitif, dan fisik. Bertambahnya usia juga meningkatkan kondisi tekanan darah baik sistolik maupun diastolic, salah satu penyakit yang berkaitan erat dengan proses penuaan ialah gangguan sirkulasi darah termasuk tekanan darah secara berlebihan dan hampir konstan pada arteri [1]. Hipertensi ialah penyakit yang paling umum di derita oleh orang tua dan kelompok lanjut usia, dikarenakan factor fisiologis manusia normal yang mengalami penurunan fungsi seiring dengan bertambahnya usia [2].

Hipertensi ditandai dengan peningkatan tekanan darah diastolic yang melebihi 90 mmHg dan peningkatan tekanan darah sistolik yang melebihi 140 mmHg, yang diamati dalam dua pengukuran berturut-turut yang dilakukan pada interval lima menit dalam kondisi tenang. Prevalensi hipertensi secara konsisten meningkat secara global, mempengaruhi sekitar 1 miliar orang atau 1 dari 4 orang dewasa, dan diproyeksikan mencapai 1,6 miliar orang pada tahun 2025 [3]. Dalam konteks Indonesia, data Riset Kesehatan Dasar (RISKESDAS) tahun 2018 [4] menunjukkan bahwa hipertensi telah mempengaruhi rata-rata 34,1 persen dari populasi. Tingginya angka kejadian hipertensi dapat disebabkan oleh dua factor intrinsic, yaitu usia dan jenis kelamin yang tidak dapat dimodifikasi, serta factor ekstrinsik yang dipengaruhi oleh pilih gaya hidup yang dapat dimodifikasi seperti konsumsi alcohol, obesitas, perilaku kurang gerak, merokok, konsumsi kopi, penggunaan pil Keluarga Berencana (KB), dan konsumsi garam yang berlebihan [1].

Pasien dengan hipertensi tidak memiliki gejala spesifik atau disebut sebagai penyakit sunyi dikarenakan para penderitanya tidak mengetahui bahwasannya mereka memiliki hipertensi sebelum melakukan pemeriksaan terhadap tekanan darah mereka, sedangkan hipertensi dapat mengakibatkan timbulnya penyakit serius hingga kompilasi sehingga dinyatakan sebagai Silent Killer [2]. Penyakit ini tergolong sebagai heterogen yang dapat menyerang seluruh kelompok umur termasuk usia produktif, akan tetapi kelompok lanjut usia memiliki risiko yang lebih tinggi, dan tidak dapat disembuhkan atau berkaitan dengan penurunan harapan hidup. Pasien dengan hipertensi umumnya akan tersadar bahwasannya mereka memiliki penyakit tersebut apabila telah merasakan berbagai gelaja yang tidak tertangani secara pribadi atau semakin parah berupa epiktasis, sesak nafas, gelisah, lemas, kelelahan, nyeri tengkuk, muntah, mual, sakit kepala hebat, hingga terjadinya penurunan kesadaran [1].

Pasien dengan hipertensi yang tidak terdiagnosis sedari dini dan tidak diobati, akan mengalami kesulitan dalam mengontrol tekanan darah secara jangka panjang, hingga memicu timbulnya peningkatan risiko komplikasi dari hipertensi dan berbagai penyakit kardiovaskular. Hipertensi telah menjadi penyebab kematian nomor satu di Indonesia, dengan dampak terburuk ialah kematian yang mana hingga saat ini diperkirakan telah menyebabkan sebanyak 7,5 miliar kematian atau 12,8 persen dari total persentase kematian yang ada [3]. Menurut World of Health Organization (WHO) tahun 2023 dalam laman [5] hipertensi ialah kondisi medis yang serius dan dapat meningkatkan risiko penyakit ginjal, otak, jantung, hingga menjadi penyebab utama terjadinya kematian dini di seluruh dunia dengan 1 dari 5 orang wanita dan 1 dari 4 orang pria. Dengan demikian, diperlukan informasi terkait pola gejala dari hipertensi agar membantu para pasien untuk menyadari sedari dini dan melakukan tindak lanjut sesegera mungkin, sekaligus sebagai pemberian pengetahuan baru terkait gelaja hipertensi di suatu layanan Kesehatan [5].

Machine learning (ML) merupakan sub bidang kecerdasan buatan (AI) yang memproses dan belajar dari data dalam jumlah besar dan menggunakan data tersebut untuk membangun model dengan mempelajari pola-pola dalam data dan apa yang diharapkan oleh model tersebut [6]. Kehadiran kecerdasan buatan dalam bidang Kesehatan telah menyebabkan peningkatan pesat dalam kualitas layanan medis. Machine learning dan deep learning merupakan dua model kecerdasan buatan yang dengan mudah dan cepat mempelajari dan memproses data untuk mendapatkan informasi yang dibutuhkan.

Identifikasi hipertensi melalui pembelajaran mesin telah dieksplorasi secara ekstensif dalam penelitian terbaru. Dalam sebuah penelitian [7], Artificial Neural Network digunakan untuk memprediksi hipertensi berdasarkan data klinis pasien mencapai tingkat akurasi model yang luar biasa yaitu 82%. Penelitian lain [8] menggunakan algoritma Naïve Bayes pada data klinis menghasilkan tingkat akurasi model yang patut dipuji sebesar 83,69%. Selanjutnya, penelitian [9] berfokus pada prediksi hipertensi menggunakan algoritma regresi logistic dan Artificial Neural Network dengan tingkat akurasi 72% berdasarkan data kuesioner. Selain itu, penelitian terpisah [10] mengembangkan model prediksi untuk mengklasifikasikan tingkat risiko hipertensi dengan menggunakan algoritma Decision Tree untuk mengidentifikasi variabel-variabel yang berpengaruh dalam memprediksi tingkat risiko hipertensi pada individu. Temuan-temuan ini menggarisbawahi efektivitas pembelajaran mesin dalam prediksi hipertensi, dengan berbagai algoritma yang menunjukkan tingkat akurasi yang menjanjikan.

Berdasarkan wawasan dari penelitian sebelumnya, penelitian ini bertujuan untuk mengidentifikasi kasus hipertensi melalui penerapan algoritma klasifikasi machine learning. Secara khusus, algoritma Naïve Bayes, Decision Tree, dan Artificial Neural Network (ANN) akan digunakan dalam penelitian ini.

1. Materials and Methods

Hipertensi merupakan ancaman kesehatan yang signifikan di Indonesia, menduduki peringkat ketiga penyebab kematian di semua kelompok umur (6,8%), setelah stroke (15,4%) dan tuberkulosis (7,5%) [11]. Dijelaskan dalam penelitian [12] sebagai gangguan sistem peredaran darah, hipertensi ditandai dengan tingkat tekanan darah yang tinggi, yang berpotensi menyebabkan komplikasi parah seperti stroke dan penyakit jantung yang fatal. Menurut Organisasi Kesehatan Dunia (WHO), hipertensi, atau "tekanan darah tinggi", didefinisikan sebagai tekanan darah sistolik >= 140 mmHg dan tekanan darah diastolik >= 90 mmHg [13]. Tekanan darah sistolik mencerminkan kontraksi jantung, sedangkan tekanan darah diastolik menandakan relaksasi jantung selama pengisian darah [13]. Memahami dan menangani hipertensi sangat penting untuk mencegah konsekuensi yang mengancam jiwa.

Hipertensi esensial atau primer, juga disebut sebagai hipertensi idiopatik, tidak memiliki penyebab yang jelas. Bentuk hipertensi ini sangat lazim, mencakup sekitar 90% kasus hipertensi [14]. Pada hipertensi primer, tidak ada penyakit renovascular, aldosteronisme, pheochromocytoma, gagal ginjal, atau penyakit lain yang mendasarinya. Faktor yang diduga menjadi penyebab hipertensi primer meliputi genetika, ras, factor lingkungan, dan gaya hidup yang tidak sehat, termasuk konsumsi alcohol dan merokok.

Penyakit jantung hipertensi adalah istilah yang luas uang mencakup kondisi jantung seperti hipertrofi ventrikel kiri, penyakit arteri coroner, aritma jantung, dan gagal jantung kongestif, yang semuanya berasal dari peningkatan tekanan darah yang kronis [15]. Meskipun akar penyebab penyakit jantung hipertensi adalah hipertensi kronis, factor-faktor yang berkontribusi terhadap hipertensi beragam [15].

**2.1 Pengolahan Data dan Penyajian Data**

Pemrosesan data melibatkan transformasi data mentah menjadi informasi yang dirangkum, biasanya disajikan dalam bentuk rangkuman angka. Pemrosesan ini dapat dilakukan secara manual atau dengan bantuan perangkat elektronik seperti kalkulator dan computer. Hasil dari pemrosesan data dapat mengambil berbagai format, termasuk gambar, grafik, table, audio, dan banyak lagi [16].

Penyajian data statistic adalah metode pengorganisasian data statistic untuk meningkatkan visibilitas dan pemahaman bagi pengguna data. Tujuannya adalah untuk memfasilitasi pemahaman dan analisis data statistic, sehingga meningkatkan akurasi dalam menarik kesimpulan dan mengambil keputusan berdasarkan data tersebut. Beberapa jenis representasi grafis yang digunakan untuk penyajian data, antara lain diagram batang, diagram lingkaran, distribusi frekuensi, histogram, polygon, dan lain-lain [17].

**2.2 Data Mining**

Data mining terdiri dari serangkaian prosedur yang bertujuan untuk mengekstraksi nilai tambahan dari sebuah dataset dalam bentuk pengetahuan yang sebelumnya tidak diketahui (Pramdhani). Dalam ranah penelitian (Yuli, 2017), sesuai dengan definisi Gartner Group, data mining melibatkan penemuan hubungan, pola, dan kebiasaan baru yang bermakna dengan menyaringdata yang luas yang disimpan dalam berbagai media penyimpanan, memanfaatkan teknologi pengenalan pola seperti teknik statistic dan matematika. Sering disinonimkan dengan *knowledge discovery in databases* (KDD), berbagai tahapan proses data mining digambarkan pada **Figure 1**.

A diagram of a data processing process

Description automatically generated

Figure 1 Knowledge Discovery in Databases (KDD) process[18]

**2.3 Selection**

Pemilihan data berkaitan dengan kebutuhan untuk mengembangkan dan memahami domain aplikasi, menangkap pengetahuan sebelumnya yang relevan, dan mengidentifikasi tujuan proses KDD dari perspektif pengguna akhir. Setelah itu, kumpulan data target yang sesuai atau subset variabel harus dipilih.

**2.4 Pre-processing**

Pra-pemrosesan mencakup pembersihan data dalam hal penanganan bidang data yang hilang / missing value, penghapusan duplikasi, serta penggabungan dan resolusi konflik karena data berasal dari sumber yang heterogen. [19] menyatakan pentingnya persiapan data dalam menghasilkan pengetahuan yang berkualitas tinggi melalui KDD. Selain itu, [20] menunjukkan bahwa 60% dari waktu yang digunakan untuk persiapan data, sedangkan penambangan itu sendiri hanya menyumbang 10% dari keseluruhan upaya.

**2.5 Transformation**

Transformasi berkaitan dengan reduksi dan proyeksi data dengan tujuan menemukan fitur yang berguna dan merepresentasikan data sesuai dengan kebutuhan tujuan yang dinyatakan dan algoritma yang dipilih. Hal ini termasuk menemukan representasi data yang tidak berubah-ubah dan menggunakan metode reduksi dimensi untuk mengurangi jumlah variabel yang dipertimbangkan secara efektif.

**2.6 Data Mining**

Data mining berhubungan dengan pencocokan tujuan KDD yang telah ditentukan dengan metode tertentu, misalnya klasifikasi, regresi, atau clustering. Hal ini mencakup pemilihan algoritma dan metode ekstraksi pola, serta pertimbangan mengenai kemampuan pengguna akhir untuk menginterpretasikan model yang dipilih dengan kemampuan prediksi dan akurasi model. Data mining yang sebenarnya kemudian dapat dilakukan pencarian pola dalam bentuk representasi tertentu atau sekumpulan representasi, seperti rule sets, trees, clusters, dll.

**2.7 Interpretation / Evaluation**

Langkah terakhir melibatkan interpretasi pola-pola ranjau dan pemeriksaan validitasnya. Hal ini dapat mencakup visualisasi pola yang ditemukan dan penilaian kegunaannya. Yang paling penting adalah bertindak berdasarkan pengetahuan yang ditemukan, misalnya mendokumentasikannya, menggunakannya secara langsung, atau mengimplementasikannya ke dalam system lain untuk digunakan lebih lanjut.

**2.8 Data**

Penelitian ini fokus pada analisis data pasien hipertensi di Puskesmas Modopuro Mojokerto selama periode enam bulan, yakni dari bulan Maret hingga Agustus 2022. Dalam menjalankan penelitian ini, peneliti sangat memperhatikan kerahasiaan pasien dengan tidak mengakses informasi rinci mengenai Riwayat medis atau kondisi Kesehatan pasien. Pendekatan ini diambil sebagai bentuk penghormatan terhadap privasi pasien dan sebagai kepatuhan terhadap etika penelitian yang sesuai dengan standar penelitian yang berlaku.

Dari dataset yang dikumpulkan selama periode tersebut, terdapat 80 fitur yang mencakup informasi mengenai pasien hipertensi dengan total 1734 data pasien. Namun, demi menjaga privasi, peneliti memilih untuk menggunakan 14 fitur dengan melibatkan 1526 data pasien. Langkah ini diambil sebagai upaya lebih lanjut dalam memastikan keamanan informasi pasien sejalan dengan prinsip-prinsip etika penelitian yang telah ditetapkan.

Dalam proses klasifikasi pasien hipertensi di Puskesmas Modopuro Mojokerto, sejumlah 14 fitur atau kriteria digunakan untuk memberikan gambaran yang komprehensif. Berikut adalah penjelasan singkat mengenai masing-masing fitur atau kriteria tersebut:

1. Jenis Kelamin: Menunjukkan jenis kelamin pasien sebagai factor potensial yang dapat memengaruhi kondisi Kesehatan.
2. Umur (Tahun): Representasi numerik dari usia pasien, memberikan gambaran tentang sejauh maan usia dapat memengaruhi risiko terkait penyakit.
3. Merokok: Menunjukkan apakah pasien merokok atau tidak, karena kebiasaan merokok dapat menjadi factor risiko yang signifikan.
4. Konsumsi alcohol: Menggambarkan apakah pasien mengonsumsi alcohol atau tidak, karena konsumsi alcohol dapat berpengaruh pada Kesehatan.
5. Kurang sayur/buah: Memberikan informasi tentang pola makan pasien terkait konsumsi sayur dan buah yang dapat mempengaruhi Kesehatan.
6. Tinggi (Cm): Tinggi badan pasien.
7. Berat Badan (Kg): Berat badan pasien.
8. IMT (Indeks Massa Tubuh): Mengukur hubungan antara berat badan dan tinggi badan pasien.
9. Hasil IMT: Memberikan kategori hasil IMT, memberikan indikasi tentang status berat badan pasien.
10. Sistole (mm): Tekanan darah sistolik pasien, mencerminkan tekanan darah saat jantung berkontraksi.
11. Diastole (Hg): Tekanan darah diastolic pasien, mencerminkan tekanan darah saat jantung beristirahat di antara kontraksi.
12. Nafas (Menit): Frekuensi pernapasan pasien per menit.
13. Detak Nadi (Menit): Frekuensi detak jantung pasien per menit.
14. Diagnosa 1: Hasil diagnose awal terkait kondisi Kesehatan pasien.

**2.9. Mean Imputation (MI)**

Metode Mean digunakan secara luas untuk menghitung data yang hilang, yang melibatkan penggantian nilai yang hilang dalam suatu variabel dengan rata-rata semua nilai yang diketahui dalam variabel tersebut [21]. Namun, metode ini memiliki kelemahan karena mengarah pada pengurangan varians variabel, mengingat bahwa nilai yang diganti seragam di semua contoh [22].

**2.10. Klasifikasi**

Klasifikasi adalah prosedur yang melibatkan pelatihan atau pembelajaran fungsi target, dilambangkan sebagai , yang memetakan setiap vector (kumpulan fitur) ke salah satu dari beberapa label kelas yang tersedia . Set atau kumpulan ini mencakup berbagai atribut, yang dapat berupa kontinu atau kategorikal. Salah satu atribut ini menentukan kelas untuk setipa record [23]. **Figure 2** mengilustrasikan model klasifikasi.

A close-up of a sign

Description automatically generated

Figure 2 Blok diagram model klasifikasi

Klasifikasi adalah metode yang digunakan untuk mengembangkan model untuk menjelaskan atau membedakan antara konsep atau kelas data. Tujuannya adalah untuk dapat memprediksi kelas suatu objek dengan label yang tidak diketahui [23].

**2.11. Naïve Bayes**

Algoritma Naïve Bayes adalah teknik klasifikasi yang diakui sebagai metode optimal dalam domain ini, yang awalnya diperkenalkan oleh Thomas Bayes. Algoritma ini memprediksi kejadian di masa depan dengan memanfaatkan data masa lalu. Naïve Bayes menggunakan fungsi statistic langsung yang didasarkan pada teorema Bayes, dengan mengansumsikan independensi fitur dari fitur lain dalam suatu kelas [24]. Ini adalah metode klasifikasi bebasis probabilitas [25].

Persamaan dari probabilitas prior:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Informasi:

*:* Jumlah data pada suatu classs

*N:* Jumlah total data

Persamaan dari teorema bayes:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Informasi:

*A :* Data Kelas

*B :* Data Hipotesis

*P(B|A) :* Probabilitas hipotesis B terhadap kondisi A

*P(B) :* Probabilitas hipotesis B

*P(A|B) :* Probabilitas A terhadap kondisi B

*P(A)* : Probabilitas A

**2.12. Decision Tree**

Pohon keputusan befungsi sebagai ilustrasi langsung dari metode klasifikasi yang dirancang untuk sekumpulan kelas yang terbatas. Dalam representasi ini, nama atribut melabeli node internal dan daun, sementara nilai atribut potensial melabeli cabang. Simpul terminal, atau daun diberi keterangan dengan kelas-kelas yang berbeda[26].

**2.13. Algoritma C4.5**

Algoritma C4.5 digunakan untuk membangun pohon keputusan, sebuah metode yang sangat efektif dan dikenal luas untuk klasifikasi dan prediksi. Algoritma ini membangun pohon keputusan dari data pelatihan, yang terdiri dari kasus atau catatan (tupel) dalam database. Setiap kasus memiliki nilai atribut yang sesuai dengan kelas, dengan atribut yang mengakomodasi data diskrit dan kontinu (numerik). Algoritma C4.5 dapat menangani kasus dimana satu atau lebih atribut tidak memiliki nilai, meskipun atribut kelas harus bertipe diskrit dan tidak boleh kosong [27]. Proses umum untuk membangun pohon keputusan menggunakan algoritma C4.5 adalah dengan memilih sebuah atribut sebagai akar, membuat cabang untuk setiap nilai atribut, membagi kasus dalam setiap cabang, dan mengulangi proses tersebut hingga semua kasus dalam sebuah cabang memiliki kelas yang sama.

**2.13. Algoritma ID3 (Iterative Dichotomiser)**

Algoritma ID3 merupakan sebuah pendekatan pembelajaran Decision Tree dasar yang dikembangkan oleh J. Ross Quinlan pada tahun 1993, melakukan eksplorasi komprehensif terhadap semua pohon keputusan yang dapat dibanyangkan [28]. Secara singkat, langkah-langkah operasional algoritma ID3 dapat diuraikan sebagai berikut:

1. Masukkan data pelatihan dan variabel
2. Hitung Entropy dan Information gain tertinggi

Menghitung nilai Entropy:

|  |  |
| --- | --- |
|  | (3) |

Informasi:

*S :* Himpunan kasus

*A :* Fitur

*n :* Jumlah partisi S

*pi :* Proporsi dari Si terhadap S

Menghitung Information Gain:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Informasi:

*S :* Himpunan

*A :* Atribut

*|Si| :* Jumlah kasus pada partisi ke-i

*|S| :* Jumlah kasus dalam S

1. Pilih variabel dengan nilai information gain tertinggi
2. Buat sebuah node yang berisi variabel yang dipilih tersebut
3. Ulangi proses perhitungan information gain hingga semua data masuk ke dalam kelas yang sama. Variabel yang sudah dipilih tidak diikutsertakan dalam perhitungan nilai information gain selanjutnya [29].

**2.14. Artificial Neural Network (ANN)**

Jaringan saraf dapat digunakan di berbagai bidang, termasuk penggalian data untuk tugas-tugas seperti klasifikasi dan prediksi. Awalnya dirancang dalam pembelajaran mesin untuk meniru aspek neurofisiologis otak manusia, jaringan saraf melibatkan penggabungan elemen komputasi sederhana (neuron) ke dalam sistem yang saling terhubung [30]. Seperti yang diuraikan dalam [31], baik algoritma neural network maupun aktivitas otak manusia terutama berfokus pada dua elemen kunci: pengetahuan yang diperoleh dari proses pembelajaran dan kekuatan koneksi antar neuron, yang disebut sebagai bobot, yang berfungsi sebagai media untuk menyimpan pengetahuan yang diperoleh. Jaringan syaraf diakui sebagai algoritma yang efektif untuk menangani data yang beragam, terutama dengan adanya gangguan (noise) [32].

**2.15. Algoritma Backpropagation**

Algoritma ini merupakan bagian dari teknik jaringan syaraf tiruan, yang secara khusus dirancang untuk meminimalisir kesalahan pada output jaringan. Backpropagation memungkinkan pelatihan jaringan untuk mencapai keseimbangan antara mengenali pola yang digunakan dalam pelatihan dan memberikan respons yang akurat terhadap pola yang mirip dengan pola yang ditemui selama pelatihan [33]. Proses pelatihan backpropagation terdiri dari tiga fase:

1. Fase fordward melibatkan perhitungan pola data pelatihan dari lapisan input ke lapisan output..
2. Back fordward terjadi Ketika setiap neuron output menerima pola input target, menyebarkan nilai kesalahan yang telah dihitung ke depan.
3. Fase modifikasi bobot diimplementasikan untuk mengurangi nilai kesalahan [34].

Jaringan syaraf tiruan yang menggunakan backpropagation mengikuti topologi multi-layer, yang terdiri dari satu lapisan input dengan , satu atau lebih lapisan tersembunyi dan satu lapisan output Y. Neuron-neuron pada setiap lapisan digambarkan sebagai lingkaran, dengan koneksi antara neuron pada satu lapisan dan neuron pada lapisan berikutnya yang dicirikan dengan bobot pada output node ()dan bobot pada hidden node (). Lapisan tersembunyi memasukkan nilai bias (b) dengan bobot yang sama dengan 1 [35].

**2.16. Evaluasi**

Dalam memodelkan proses klasifikasi, perlu dilakukan evaluasi terhadap performa sistem untuk mengukur seberapa baik metode yag digunakan [28]. Metode yang biasa digunakan dalam mengevaluasi sistem adalah confusion matrix. Confusion matrix merupakan salah satu metode evaluasi perhitungan tingkat akurasi, hasil precision, recall, dan f-measure dari algoritma yang digunakan dalam penelitian dan diukur dari hasil data testing yang telah diprediksi [29].

1. Akurasi adalah nilai keefektifan dari keseluruhan hasil proses klasifikasi, dengan rumus sebagai berikut:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

1. Precision adalah hasil dari presentase label data klasifikasi positif, dengan rumus sebagai berikut:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

1. Recall adalah hasil dari keefektifan proses klasifikasi untuk identifikasi label positif, dengan rumus sebagai berikut:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

1. F-measure adalah hasil dari rata-rata recall dan precision, dimana kisaran f-measure itu sendiri adalah 0-1.

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

Informasi:

*TP*: jumlah data yang diprediksi dengan benar positif,

*TN*: jumlah data dengan kelas asli positif tetapi hasil prediksi negatif,

*FN*: jumlah data yang diprediksi dengan benar negatif,

*FP*: jumlah data dengan kelas asli negatif tetapi hasil prediksi positif.

1. RESULT AND DISCUSSION

3.1. Pengumpulan Data

Data yang digunakan dalam proses ini adalah data hipertensi di Puskesmas Modopuro Mojokerto yang terdiri dari 1526 data pasien dan 14 atribut, yaitu Jenis Kelamin, Umur (Tahun), Merokok, Konsumsi Alkohol, Kurang Sayur/Buah, Tinggi (Cm), Berat Badan (Kg), IMT (Indeks Massa Tubuh), Hasil IMT, Sistole (mm), Diastole (Hg), Nafas (Menit), Detak Nadi (Menit), termasuk kelas Diagnosa 1 yang terdiri dari 2 kelas dengan rincian 1429 pasien tekena Essential (primary) hypertension (I10) dan 97 pasien terkena Hypertensive heart disease (I11). Penjelasan data diatas dapat dilihat pada Tabel 1 di bawah ini.

Table 1 Dataset Hypertensi

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Jenis Kelamin** | **Umur (Tahun)** | **Merokok** | **Konsumsi Alkhohol** | **Kurang Sayur/Buah** | **Tinggi (Cm)** | **Berat Badan (Kg)** | **IMT** |
| 1. | 0 | 63 | 0 | 0 | 0 | 147 | 43 | 19.90 |
| 2. | 0 | 55 | 0 | 0 | 0 | 145 | 45 | 21.40 |
| 3. | 0 | 70 | 0 | 0 | 0 | 146 | 45 | 21.11 |
| 4. | 0 | 66 | 0 | 0 | 0 | 165 | 55 | 20.20 |
| 5. | 0 | 65 | 0 | 0 | 0 | 158 | 56 | 22.43 |
| 6. | 0 | 76 | 0 | 0 | 0 | 158 | 56 | 22.43 |
| 7. | 0 | 60 | 0 | 0 | 0 | 148 | 46 | 21.00 |
| 8. | 1 | 58 | 0 | 0 | 0 | 158 | 56 | 22.43 |
| 9. | 1 | 79 | 0 | 0 | 0 | 160 | 59 | 23.05 |
| 10. | 1 | 69 | 0 | 0 | 0 | 166 | 60 | 21.77 |
| 11. | 0 | 60 | 0 | 0 | 0 | 156 | 54 | 22.19 |
| 12. | 0 | 49 | 0 | 0 | 0 | 158 | 56 | 22.43 |
| 13. | 0 | 56 | 0 | 0 | 0 | 158 | 56 | 22.43 |
| 14. | 0 | 52 | 0 | 0 | 0 | 163 | 76 | 28.60 |
| 15. | 0 | 62 | 0 | 0 | 0 | 159 | 56 | 22.15 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Hasil IMT** | **Sistole (mm)** | **Diastole (Hg)** | **Nafas (Menit** | **Detak Nadi (Menit)** | **Diagnosa 1** |
| 1. | 2.0 | 180 | 80 | 18 | 88 | 0 |
| 2. | 2.0 | 170 | 80 | 18 | 88 | 0 |
| 3. | 2.0 | 170 | 80 | 18 | 88 | 0 |
| 4. | 2.0 | 180 | 80 | 18 | 88 | 0 |
| 5. | 2.0 | 130 | 80 | 18 | 80 | 1 |
| 6. | 2.0 | 152 | 82 | 18 | 80 | 0 |
| 7. | 2.0 | 164 | 85 | 18 | 80 | 0 |
| 8. | 2.0 | 148 | 90 | 18 | 80 | 0 |
| 9. | 2.0 | 170 | 80 | 18 | 88 | 0 |
| 10. | 2.0 | 180 | 80 | 18 | 88 | 0 |
| 11. | 2.0 | 165 | 90 | 20 | 81 | 0 |
| 12. | 2.0 | 130 | 80 | 18 | 80 | 1 |
| 13. | 2.0 | 183 | 80 | 18 | 80 | 0 |
| 14. | 3.0 | 150 | 90 | 20 | 80 | 0 |
| 15. | 2.0 | 198 | 114 | 18 | 80 | 0 |

**3.2. Analisis**

Untuk menjelaskan tahapan atau prosedur investigasi dalam menjawab tujuan penelitian, peneliti menggambarkan langkah-langkah penelitian:

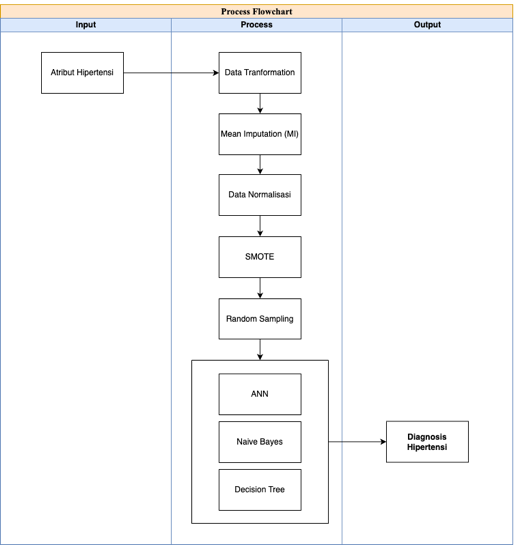


Figure 3 Process Flowchart

1. Data Input Process

Pada proses ini, data inputan yang digunakan untuk proses identifikasi data pasien hipertensi adalah 14 fitur.

1. Data Pre-processing

Pada tahap ini dilakukan proses cleaning data untuk menghapus data yang tidak mempunyai relasi atau tidak diperlukan saat proses identifikasi, normalisasi data untuk menormalisasi value agar berupa range dari 0-1 dengan menggunakan MinMaxScaler.

1. Data Balancing Process

Pada tahap ini dilakukan proses SMOTE untuk menyeimbangkan data agar jumlah data pada masing-masing kelas seimbang dengan menggunakan k=5. Berikut perbandingan data sebelum dan sesudah SMOTE.

|  |  |  |
| --- | --- | --- |
| **Diagnosa 1** | **Sebelum SMOTE** | **Sesudah SMOTE** |
| 0 | 1429 | 1429 |
| 1 | 97 | 1429 |

1. Data Sharing Process

Metode random sampling digunakan untuk membagi data latih dan data uji dengan pemilihan scenario 90:10, 80:20, dan 70:30. Data training/latih digunakan dalam membangun model dengan jumlah data tertentu, sedangkan data test/uji adalah sisa data yang tidak digunakan pada saat training untuk menguji performa model yang telah dilatih.

1. Classification Process

Pada tahap ini dilakukan proses pembelajaran untuk mendapatkan model klasifikasi dengan menggunakan beberapa metode yang berbeda, yaitu Artificial Neural Network (ANN), Naïve Bayes, dan Decision Tree. Model-model klasifikasi yang berbeda tersebut akan dibandingkan hasil akurasinya untuk mendapatkan hasil klasifikasi yang optimal.

1. Output

Setelah seluruh proses dijalankan, maka akan menghasilkan output berupa prediksi kelas dari Diagnosis 1 berdasarkan pemodelan dengan metode yang diusulkan dalam penelitian ini.

**3.3. Artificial Neural Network (ANN)**

Penggunaan algoritma Artificial Neural Network (ANN) dilakukan dengan membuat uji coba untuk menarik kesimpulan yang tepat setelah proses penelitian dilakukan. Skenario uji coba yang dilakukan pada penelitian ini dijelaskan pada tabel .

Table 2 Testing Skenario

|  |  |  |
| --- | --- | --- |
| **No** | **Skenario Uji Coba** | **Parameter** |
| 1. | Random Sampling | 90:10, 80:20, 70:30 |
| 2. | Neuron | (8, 8, 8), (16, 16, 16), (8, 16,8) |
| 3. | Optimasi | adam, sgd |
| 4. | Fungsi Aktivasi | relu, tanh |
| 5. | Learning Rate | constant, adaptive |
| 6. | Max Iter | 25, 50, 100, 1000 |

Table 3 Parameter Terbaik dari Grid Search Results

|  |  |  |
| --- | --- | --- |
| **No** | **Skenario Uji Coba** | **Parameter** |
| 1. | Neuron | (8, 8, 8) |
| 2. | Optimasi | adam |
| 3. | Fungsi Aktivasi | relu |
| 4. | Learning Rate | constant |
| 5. | Max Iter | 25 |

Berdasarkan parameter-parameter pada skenario pengujian, kemudian dicari parameter terbaik dengan mengoptimasi parameter-parameter tersebut menggunakan Grid Search. pengujian menggunakan fungsi aktivasi relu dengan neuron (8,8,8) dan menggunakan optimasi sgd. Percobaan menggunakan random sampling yang berbeda yaitu 90:10, 80:20, dan 70:30 dan learning rate constant dengan banyak iterasi sebanyak 25.

Table 4 Hasil Evaluasi Menggunakan Metode ANN dalam persen

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RS** | **Dengan SMOTE** | | | | **Tanpa SMOTE** | | | |
|  | **Acc** | **Prec** | **Rec** | **F1-S** | **Acc** | **Prec** | **Rec** | **F1-S** |
| **9:1** | 74.8 | 76.6 | 74.8 | 74.8 | 93.4 | 100 | 93.4 | 90.3 |
| **8:2** | 74.5 | 72.3 | 74.5 | 74.5 | 94.1 | 100 | 94.1 | 91.3 |
| **7:3** | 76.6 | 72.9 | 76.6 | 76.5 | 93.9 | 100 | 93.9 | 90.9 |
| **Avg** | 75.3 | 73.9 | 75.3 | 75.2 | **93.8** | 100 | 93.8 | 90.8 |
| **RT** | 0.04(s) | | | | 0.05(s) | | | |

Acc : Accuracy

Prec : Precission

Rec : Recall

F1-S : F1-Score

Avg : Average

RT : Running Time

RS : Random Sampling.

Tabel 4 menunjukkan bahwa akurasi rata-rata adalah 75,3% dengan waktu komputasi 0.04 detik Ketika metode SMOTE diterapkan, sedangkan tanpa SMOTE akurasi rata-rata adalah 93,8% dengan waktu komputasi 0.05 detik.

Kelebihan Artificial Neural Network (ANN) meliputi kemampuan pemrosesan parallel, penyimpanan data di seluruh jaringan, kemampuan menghasilkan output dengan data yang tidak memada, distribusi memori dan toleransi terhadap kesalahan. Sedangkan kelemahan dari Artificial Neural Network (ANN) adalah ketidakadaan pedoman khusus untuk menentukan struktur perilaku jaringan yang tidak dapat dijelaskan, ketergantungan pada hardware, kesulitan menyampaikan masalah ke jaringan dan durasi jaringan yang tidak diketahui.

**3.4. Naive Bayes**

Hasil yang diperoleh dari penerapan dataset hipertensi menggunakan algoritma Naive Bayes, baik dengan implementasi metode SMOTE maupun tanpa metode SMOTE, akan diuji dengan menggunakan rasio pengambilan sampel acak yang berbeda, yaitu 90:10, 80:20, dan 79:30.

Table 5 Hasil Evaluasi Menggunakan Metode Naive Bayes dalam persen

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RS** | **Dengan SMOTE** | | | | **Tanpa SMOTE** | | | |
|  | **Acc** | **Prec** | **Rec** | **F1-S** | **Acc** | **Prec** | **Rec** | **F1-S** |
| **9:1** | 73.6 | 70 | 73.6 | 73.5 | 90.3 | 26.4 | 90.3 | 90.8 |
| **8:2** | 72.5 | 68.4 | 72.5 | 72.3 | 90.2 | 26.9 | 90.2 | 91 |
| **7:3** | 74.1 | 70.5 | 74.1 | 74 | 90.4 | 26.4 | 90.4 | 90.8 |
| **Avg** | 73.4 | 69.6 | 73.4 | 73.3 | **90.3** | 26.6 | 90.3 | 90.9 |
| **RT** | 0 (s) | | | | 0.01(s) | | | |

Tabel 5 menunjukkan bahwa akurasi rata-rata adalah 73,4% dengan waktu komputasi 0 detik Ketika metode SMOTE diterapkan, sedangkan tanpa SMOTE akurasi rata-rata adalah 90,3% dengan waktu komputasi 0.01 detik.

Naïve Bayes menawarkan keuntungan karena dapat diterapkan pada data kuantitatif dan kualitatif, membutuhkan pelatihan data yang minimal, dapat mengabaikan nilai yang hilang dalam perhitungan, dan bekerja dengan cepat dan efisien, sehingga mudah dipahami. Di sisi lain, kelemahannya terletak pada asumsi independensi di antara variabel, yang menyebabkan berkurangnya akurasi karena adanya korelasi yang umum di antara variabel.

**3.5. Decision Tree**

Hasil yang diperoleh dari penerapan dataset hipertensi menggunakan algoritma Decision Tree, baik dengan implementasi metode SMOTE maupun tanpa metode SMOTE, akan diuji dengan menggunakan rasio pengambilan sampel acak yang berbeda, yaitu 90:10, 80:20, dan 79:30.

Table 6 Hasil Evaluasi Menggunakan Metode Decision Tree dalam persen

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RS** | **Dengan SMOTE** | | | | **Tanpa SMOTE** | | | |
|  | **Acc** | **Prec** | **Rec** | **F1-S** | **Acc** | **Prec** | **Rec** | **F1-S** |
| **9:1** | 94 | 91.9 | 94 | 94 | 89.5 | 20 | 89.5 | 89.5 |
| **8:2** | 93.3 | 91.2 | 93.3 | 93.3 | 91.8 | 29.4 | 91.8 | 91.7 |
| **7:3** | 91.5 | 89.9 | 91.5 | 91.5 | 90.8 | 25 | 90.8 | 90.8 |
| **Avg** | **92.9** | 91 | 92.9 | 92.9 | 90.7 | 24.8 | 90.7 | 90.6 |
| **RT** | 0.01 (s) | | | | 0.01(s) | | | |

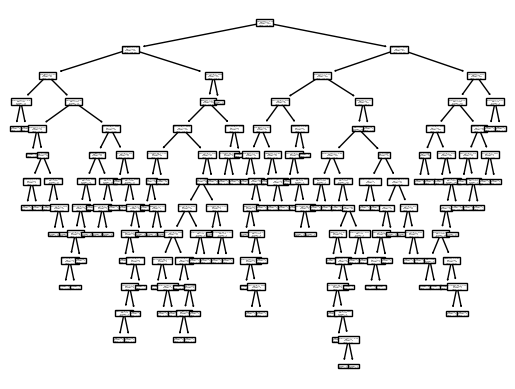


Figure 4 Plot Decision Tree

Tabel 3 menunjukkan bahwa akurasi rata-rata adalah 92,9% dengan waktu komputasi 0,01 detik Ketika metode SMOTE diterapkan, sedangkan tanpa SMOTE akurasi rata-rata adalah 90,7% dengan waktu komputasi 0.01 detik.

Kelebihan Decision Tree terletak pada fleksibilitasnya, yang memungkinkan peningkatan kualitas keputusan. Sebaliknya, kelemahannya muncul Ketika berhadapan dengan data yang menampilkan sejumlah besar kelas dan kriteria yang mengarah pada potensi tumpeng tindih.

# Conclusion

Berdasarkan analisis dan pembahasan yang dilakukan terhadap data training dan testing dengan menggunakan rasio pengambilan sampel secara acak yaitu 90:10, 80:20, dan 70:30, dimana terdapat 1526 data pasien dan 14 atribut yang meliputi Jenis Kelamin, Umur (Tahun), Merokok, Konsumsi Alkohol, Kurang Sayur/Buah, Tinggi (Cm), Berat Badan (Kg), IMT (Indeks Massa Tubuh), Hasil IMT, Sistole (mm), Diastole (Hg), Nafas (Menit), Detak Nadi (Menit), serta kelas Diagnosa 1 yang terdiri dari 2 kategori. Secara spesifik, 1429 pasien tekena hipertensi essensial (primer) (I10), dan 97 pasien terkena penyakit jantung hipertensi (I11), maka diperoleh kesimpulan bahwa:

1. Hasil pengujian menunjukkan bahwa penggunaan SMOTE meningkatkan kinerja, meskipun tanpa peningkatan akurasi yang signifikan jika dibandingkan dengan menerapkan dataset tanpa SMOTE dalam fase pengujian secara keseluruhan, hal ini terlihat dari akurasi yang diberikan.
2. Dari hasil penelitian ini dapat disimpulkan bahwa metode Artificial Neural Network (ANN) sangat mampu digunakan untuk identifikasi penyakit hipertensi.
3. Berdasarkan hasil pengujian untuk mendapatkan nilai akurasi, metode Artificial Neural Network (ANN) lebih unggul dibandingkan metode lainnya yaitu mendapatkan rata-rata akurasi 93.8% dibandingkan dengan Naïve Bayes 90.3%, dan Decision Tree 90.7% dengan scenario tanpa SMOTE.

Berdasarkan kesimpulan tersebut, penulis memberikan saran untuk pengembangan dan penelitian lebih lanjut, yaitu:

1. Penelitian selanjutnya diharapkan dapat menambahkan jumlah data dan fitur yang digunakan untuk membangun model klasifikasi, hal ini tentunya dapat mendukung performa model menjadi lebih baik lagi dengan memperhatikan jumlah missing value dan class balance pada dataset yang digunakan.
2. Penelitian selanjutnya diharapkan dapat menambahkan beberapa factor sesuai literatur seperti Riwayat keluarga dan factor risiko seperti merokok dan penggunaan alcohol serta pengkategorian obat-obatan sebagai factor lainnya.

# acknowledgment

Penelitian ini mengucapkan terima kasih atas dukungan dari Universitas Trunojoyo Madura dan Puskesmas Modopuro Mojokerto.

**REFERENCES**

[1] A. Murwani and S. Anggoro, “The Effect of Murottal Therapy on Blood Pressure in Elderly Patients With Hypertension at Jambidan Posyandu, Banguntapan 1 Community Health Center,” *JOURNAL OF NONCOMMUNICABLE DISEASES*, vol. 3, no. 1, pp. 8–15, 2023.

[2] S. Oparil *et al.*, “Hypertension,” *Nat Rev Dis Primers*, vol. 4, Mar. 2018, doi: 10.1038/nrdp.2018.14.

[3] G. Sinta Septiyawati, W. Hary Cahyati, and E. Raffy Rustiana, “Incidence of Hypertension of 30-50 Years Old in the Salatiga City Health Center,” *Public Health Perspectives Journal Gabrial Sinta Septiyawati*, vol. 6, no. 1, pp. 53–62, 2021, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/phpj

[4] Kementerian Kesehatan RI, “HASIL UTAMA RISKESDAS 2018,” 2018.

[5] World of Health Organization (WHO), “Hypertension,” https://www.who.int/health-topics/hypertension#tab=tab\_1.

[6] N. Chamidah, M. Mega Santoni, and N. Matondang, “Pengaruh Oversampling pada Klasifikasi Hipertensi dengan Algoritma Naïve Bayes, Decision Tree, dan Artificial Neural Network (ANN),” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 4, pp. 635–641, 2017.

[7] D. LaFreniere, F. Zulkernine, D. Barber, and K. Martin, “Using machine learning to predict hypertension from a clinical dataset,” in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1–7. doi: 10.1109/SSCI.2016.7849886.

[8] B. Afeni, I. Aruleba, and I. Oloyede, “Hypertension Prediction System Using Naive Bayes Classifier,” *Journal of Advances in Mathematics and Computer Science*, vol. 24, pp. 1–11, Nov. 2017, doi: 10.9734/JAMCS/2017/35610.

[9] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, “Predicting Hypertension without Measurement: A Non-invasive, Questionnaire-based Approach,” *Expert Syst Appl*, vol. 42, Nov. 2015, doi: 10.1016/j.eswa.2015.06.012.

[10] P. Idowu, “Predictive Model for the Classification of Hypertension Risk Using Decision Trees Algorithm,” *American Journal of Mathematical and Computer Modelling*, vol. 2, pp. 48–59, Nov. 2017, doi: 10.11648/j.ajmcm.20170202.12.

[11] Y. Anisa, P. Bangun, and U. Sinulingga, “Ajian Faktor Penyebab Penderita Hipertensi dengan Menggunakan Analisis Faktor di Kotamadya Medan (Studi Kasus : Rsup H. Adam Malik Medan),” *Saintia Matematika*, vol. 2, no. 4, pp. 333–343, 2014.

[12] K. Khotimah, “Stres sebagai Faktor Terjadinya Peningkatan Tekanan Darah pada Penderita Hipertensi,” *Jurnal EduHealth*, vol. 3, no. 2, 2013.

[13] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, “Badan Pengembangan dan Pembinaan Bahasa.”

[14] Tim Bumi Medika, *Berdamai dengan hipertensi*, Cetakan pertama. 2017.

[15] K. Riaz, “Hypertensive Heart Disease,” Wright State University.

[16] G. Nurvinda K and A. Widya Davita, “Ketahui Pengertian Pengolahan Data dan Metode yang Digunakan,” Dqlab.id.

[17] Dqlab.id, “Simak Contoh Visualisasi dalam Jenis Data Statistik,” Dqlab.id.

[18] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The KDD Process for Extracting Useful Knowledge from Volumes of Data,” *Commun. ACM*, vol. 39, no. 11, pp. 27–34, Nov. 1996, doi: 10.1145/240455.240464.

[19] L. Soibelman and H. Kim, “Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases,” *Journal of Computing in Civil Engineering*, vol. 16, pp. 39–48, 2002, [Online]. Available: https://api.semanticscholar.org/CorpusID:17973437

[20] P. B. Cabena, P. O. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi, “Discovering Data Mining: From Concept to Implementation,” 1997. [Online]. Available: https://api.semanticscholar.org/CorpusID:15453639

[21] E. Acuña and C. Rodriguez, “The Treatment of Missing Values and its Effect on Classifier Accuracy,” in *Classification, Clustering, and Data Mining Applications*, D. Banks, F. R. McMorris, P. Arabie, and W. Gaul, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 639–647.

[22] J. Graham, *Missing data: Analysis and design. New York, NY: Springer*. 2012. doi: 10.1007/978-1-4614-4018-5.

[23] M. Aziz Muslim *et al.*, *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program komputer*. 2019.

[24] L. Wilianto, T. Hendro Pudjiantoro, and F. Rakhmat Umbara, “Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat,” *Prosiding SNATIF*, 2017.

[25] J. Ling, I. Putu Eka N. Kencana, and T. Bagus Oka, “Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square,” *E-Jurnal Matematika*, vol. 3, no. 3, pp. 92–99, 2014.

[26] J. Eska, “Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5,” *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 2, no. 2, pp. 9–13, Apr. 2016, doi: 10.31227/osf.io/x6svc.

[27] I. Carolina and K. Ramanda, “Klasifikasi Kelahiran Prematur Menggunakan Algoritma C4.5,” *Seminar Nasional Teknologi*, 2018.

[28] M. Ture, F. Tokatli, and I. Kurt, “Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients,” *Expert Syst Appl*, vol. 36, no. 2, Part 1, pp. 2017–2026, 2009, doi: https://doi.org/10.1016/j.eswa.2007.12.002.

[29] S. Defiyanti and D. L. Crispina Pardede, “Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi Spam-Mail.” [Online]. Available: http://www.ics.uci.edu/

[30] A. M. Siregar and H. H. H, “Implementasi Algoritma Neural Network untuk Mendukung Keputusan di Desa Tamanmekar,” *Petir*, vol. 13, no. 1, 2020, doi: 10.33322/petir.v13i1.768.

[31] F. Gorunescu, *Data Mining: Concepts, models and techniques*. 2011.

[32] L. Daniel T, “Discovering Knowledge in Data: An Introduction to Data Mining,” in *Discovering Knowledge in Data*, 2004, pp. 128–146. doi: https://doi.org/10.1002/0471687545.ch7.

[33] A. Wanto, “Penerapan Jaringan Saraf Tiruan Dalam Memprediksi Jumlah Kemiskinan Pada Kabupaten/Kota Di Provinsi Riau,” *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, vol. 5, Feb. 2018, doi: 10.20527/klik.v5i1.129.

[34] D. Monika, A. Ahmad, S. Wardani, and S. Solikhun, “Model Jaringan Syaraf Tiruan Dalam Memprediksi Ketersediaan Cabai Berdasarkan Provinsi,” *Teknika*, vol. 8, pp. 17–24, Jun. 2019, doi: 10.34148/teknika.v8i1.140.

[35] A. Ahmad, P. Putri, W. Alifah, and S. Solikhun, “Analisis Jaringan Syaraf Tiruan Metode Backpropogation Dalam Memprediksi Ketersediaan Komoditas Beras Berdasarkan Provinsi Di Indonesia,” *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, vol. 2, pp. 105–118, Oct. 2019, doi: 10.31598/jurnalresistor.v2i2.358.

**LAMPIRAN**

Berikut penjelasan proyek identifikasi hipertensi menggunakan Artificial Neural Network (ANN), terlampir preview data awal yang digunakan sebelum dilakukan Transformasi maupun Normalisasi.



Figure 5 Dataset hipertensi dengan Seleksi Fitur

Setelah dilakukan seleksi fitur, selanjutnya adalah men Transformasi data dengan value string dengan format value integer.



Figure 6 Dataset hipertensi dengan Transformasi

Setelah dilakukan Transformasi data maka tahap selanjutnya adalah normalisasi data agar value antara 0-1.

A screenshot of a graph

Description automatically generated

Figure 7 Dataset hipertensi dengan Normalisasi

Selanjutnya adalah dilakukan SMOTE untuk mem balancing kan kelas minoritas agar seimbang dengan kelas mayoritas.

A black rectangular object with white text

Description automatically generated

Figure 8 Dataset hipertensi dengan SMOTE

Tahap selanjutnya adalah pembagian data train dan data test dengan uji coba 90:10 dengan data normalisasi dan SMOTE. Dan dilakukan percobaan menggunakan Artificial Neural Network (ANN) dengan menggunakan GridSearch untuk mendapatkan parameter terbaik seperti Neuron, Optimasi, Fungsi Aktivasi, Learning Rate dan Max Iter/Epoch.

A screen shot of a computer

Description automatically generated

Figure 9 Parameter terbaik untuk algoritma ANN

Dari **Figure 9** didapatkan bahwasannya parameter terbaik berada di Neuron (8,8,8), Optimasi adam, Fungsi Aktivasi relu, Learning Rate constant dan Max Iter/Epoch 25 memperoleh akuasi 0.93 atau 93%. Selanjutnya dari parameter tersebut dimasukkan kedalam algoritma ANN untuk pengujian dengan data tanpa SMOTE dan dengan SMOTE.

A screenshot of a black screen

Description automatically generated

Figure 10 hasil tanpa SMOTE dan dengan SMOTE