|  |  |
| --- | --- |
|  | **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 lebih dari 90 mmHg dan peningkatan tekanan darah sistolik lebih dari 140 mmHg pada dua pengukuran interval selama lima menit dalam kondisi tenang. Hipertensi telah meningkat pada setiap tahunnya di seluruh lingkup global dengan sebanyak 1 miliar jiwa di dunia atau 1 dari 4 orang dewasa menjadi penderita hipertensi dan diperkirakan akan meningkat menjadi 1,6 miliar jiwa pada tahun 2025 mendatang [3]. Adapun di Indonesia berdasarkan Riset Kesehatan Dasar (RISKESDAS) tahun 2018 dalam [4] yang mengemukakan bahwasannya hipertensi di Indonesia telah menjangkit hingga rata-rata 34,1 persen jiwa di tahun 2018. Tingginya insiden hipertensi tersebut disebabkan oleh dua jenis factor internal atau genetika yang tidak dapat diubah berupa usia dan jenis kelamin, kemudian factor eksternal yang pengaruhi oleh gaya hidup yang dapat diubah seperti halnya konsumsi minuman beralkohol, obesitas, kurangnya aktivitas fisik dan olahraga, kebiasaan merokok, konsumsi kopi, penggunaan pil Keluarga Berencana (KB), dan konsumsi garam secara 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.

Essential (primary) hypertension (disebut juga hipertensi idiopatik) sampai saat ini belum diketahui penyebabnya. Hipertensi ini paling sering banyak terjadi, sekitar 90% penderita hipertensi adalah hipertensi ini [14]. Pada hipertensi primer tidak ditemukan penyakit renovaskuler, aldoteronism, pheochromocytoma, gagal ginjal, dan penyakit lainnya. Genetik dan ras merupakan bagian yang diduga menjadi penyebab hipertensi primer, termasuk juga factor lain seperti lingkungan dan gaya hidup tidak sehat antara lain mengkonsumsi alcohol dan merokok.

Hypertensive heart disease adalah suatu istilah yang digunakan secara umum untuk penyakit jantung seperti hipertropi ventrikel kiri, penyakit arteri koroner, aritmia jantung, dan gagal jantung kongestif yang disebabkan oleh efek peninggian tekanan darah kronis [15]. Penyebab dari hypertensive heart disease adalah hipertensi kronis, akan tetapi penyebab dari hipertensi sangat bervariasi [15].

**2.1 Pengolahan Data dan Penyajian Data**

Pengolahan data adalah merubah data mentah untuk memperoleh keterangan-keterangan ringkasan yang berupa angka-angka ringkasan. Pengolahan data dapat dilakukan dengan manual, maupun dengan alat-alat elektronik (kalkulator, computer). Output dari pengolahan data dapat diperoleh dalam berbagai bentuk, seperti gambar, grafik, table, audio, dan lain sebagainya [16].

Penyajian data statistic adalah suatu bentuk penataan data statistic agar data lebih mudah dipandang dan mudah dipahami oleh pengguna data. Tujuannya adalah agar data statistic mudah dimengerti, mudah dianalisis, sehingga proses pengambilan kesimpulan dan keputusan berdasarkan data menjadi lebih akurat. Ada beberapa jenis penyajian data dalam bentuk grafik mulai dari grafik batang, grafik lingkaran, distribusi frekuensi, histogram, polygon, dan lain sebagainya [17].

**2.2 Data Mining**

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui (Pramadhani). Sedangkan dalam penelitian (Yuli, 2017), menurut Gartner Group, data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah Sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistic dan matematika. Istilah data mining sering disebut juga dengan knowledge discovery in database (KDD), berikut tahapan-tahapan dalam proses data mining yang ditunjukkan 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)**

Mean merupakan salah satu metode imputasi yang paling umum digunakan. Imputasi dengan metode Mean mengisi missing data dalam suatu variabel dengan rata-rata dari semua nilai yang diketahui pada suatu variabel [21]. Imputasi dengan metode Mean memiliki kelemahan yaitu mengurangi varians pada variabel [22], karena nilai yang diisikan adalah sama untuk setiap variabel.

**2.10. Klasifikasi**

Klasifikasi dapat didefinisikan sebagai suatu proses yang melakukan pelatihan/pembelajaran terhadap fungsi target yang memetakan setiap vector (set fitur) ke dalam satu dari sejumah label kelas yang tersedia. Di dalam klasifikasi diberikan sejumlah record yang dinamakan training set, yang terdiri dari beberapa atribut, atribut dapat berupa kontinyu ataupun kategoris, salah satu atribut menunjukkan kelas untuk record [23]. Model klasifikasi dapat dilihat pada **Figure 2**.

A close-up of a sign

Description automatically generated

Figure 2 Blok diagram model klasifikasi

Klasifikasi merupakan teknik yang digunakan untuk menemukan model agar dapat menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui [23].

**2.11. Naïve Bayes**

Algoritma Naïve Bayes merupakan salah satu algoritma yang termasuk kedalam klasifikasi. Metode ini merupakan metode terbaik dalam pengklasifikasian yang dikemukakan oleh Thomas Bayes dengan memprediksi peluang di masa depan berdasarkan data di masa sebelumnya.

Naïve Bayes menerapkan fungsi statistic sedeerhana berdasarkan teorema bayes dengan asumsi keberadaan dari suatu fitur tertentu terhadap suatu kelas yang tidak berhubungan dengan fitur lainnya [24]. Naïve Bayes merupakan suatu metode klasifikasi yang menggunakan perhitungan 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 merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akan ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda [26].

**2.13. Algoritma C4.5**

Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membentuk decision tree (pohon keputusan). Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Algoritma C4.5 menkonstruksi pohon keputusan dari data pelatihan, yang berupa kasus-kasus atau record-record (tupel) dalam basis data. Setiap kasus berisikan nilai dari atribut-atribut untuk sebuah kelas. Setiap atribut dapat berisi data diskret dan kontinyu (numerik). Algoritma C4.5 juga menangani kasus yang tidak memiliki nilai untuk sebuah atau lebih atribut. Akan tetapi, atribut kelas hanya bertipe diskret dan tidak boleh kosong [27]. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing-masing nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

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

Algoritma ID3 adalah algoritma pembelajaran Decision Tree yang sederhana dikembangkan oleh J. Ross Quinlan pada tahun 1993. Algoritma ID3 melakukan prosedur pencarian secara menyeluruh pada semua kemungkinan pohon keputusan [28]. Secara ringkas, Langkah kerja algoritma ID3 adalah sebagai berikut:

1. Input data training dan variabel.
2. Hitung Entropy dan Information Gain dari setiap variabel.

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 yang memiliki nilai information gain terbesar.
2. Bentuk simpul yang berisi variabel tersebut.
3. Ulangi proses perhitungan information gain yang akan terus dilaksanakan samapi semua data telah masuk dalam kelas yang sama. Variabel yang telah dipilih tidak diikutkan lagi dalam perhitungan nilai information gain [29].

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

Jaringan saraf atau artificial neural networks dapat digunakan untuk berbagai tujuan termasuk klasifikasi dan prediksi dalam penambangan data. Jaringan saraf pada awalnya dibangun dalam pembelajaran mesin untuk mencoba meniru neurofisiologi otak manusia melalui kombinasi elemen komputasi sederhana (neuron) ke dalam sistem yang saling berhubungan [30]. Menurut [31], aktivitas algoritma neural network dan otak manusia pada dasarnya terkonsentrasi dalam dua aspek yaitu pengetahuan didapatkan dari proses pembelajaran dan intensitas koneksi antar neuron dikenal sebagai bobot sebagai media yang dapat menyimpan pengetahuan yang diperoleh. Neural network dikenal sebagai algoritma yang baik untuk menangani berbagai data yang memiliki noise [32].

**2.15. Algoritma Backpropagation**

Algoritma ini merupakan salah satu teknik dari artificial neural network. Backpropagation dapat meminimalisir kesalahan dalam hasil luaran yang dihasilkan oleh jaringan. Algoritma ini dapat melatih jaringan agar menghasilkan keseimbangan antara kemampuan jaringan dalam membaca pola yang digunakan pada saat proses pelatihan dan kemampuan dalam menghasilkan respon yang tepat pada pola yang serupa dengan pola yang digunakan pada proses pelatihan [33]. Terdapat tiga fase dalam pelatihan backpropagation antara lain yaitu:

1. Fase fordward dimana pada fase ini pola data latih dihitung mulai dari lapisan input hingga lapisan output.
2. Back fordward yaitu saat masing-masing neuron output menerima target pola input agar terhitung nilai kesalahan yang dipropagasikan mudur.
3. Fase modifikasi bobot yang bermanfaat untuk menurunkan nilai kesalahan muncul [34].

Artificial neural network dengan backpropagation merupakan jenis topologi multi layer dengan komposisi satu layer masukan dengan , satu atau lebih lapisan hidden layer dan satu lapis output yaitu Y. Masing-masing lapisan mempunyai neuron-neuron yang dimodelkan dengan lingkaran. Antara neuron pada satu layer dengan satu neuron layer berikutnya terhubung dengan koneksi yang memiliki bobot ke simpul output dan bobot ke simpul hidden . Pada hidden layer memiliki nilai bias (b) yang memiliki bobot 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), Diagnosa 1. Pada data yang digunakan terdapat missing value yang akan diatasi dengan mengimplementasikan imputasi mean. Nilai missing value pada dataset dapat dilihat pada Tabel 3 di bawah ini.

Table 1 Jumlah Missing Value pada Setiap Atribut

|  |  |
| --- | --- |
| **Atribut** | **Jumlah Missing Value** |
| Jenis Kelamin | 0 |
| Umur (Tahun) | 0 |
| Merokok | 0 |
| Konsumsi Alkohol | 0 |
| Kurang Sayur/Buah | 0 |
| Tinggi (Cm) | 0 |
| Berat Badan (Kg) | 0 |
| IMT (Indeks Massa Tubuh) | 1 |
| Hasil IMT | 1 |
| Sistole (mm) | 0 |
| Diastole (Hg) | 0 |
| Nafas (Menit) | 0 |
| Detak Nadi (Menit) | 0 |
| Diagnosa 1 | 0 |

**3.2. Analisis**

Untuk memperjelas tentang tahapan atau Langkah penelitian dalam menjawab tujuan dari penelitian ini, peneliti membuat tahap-tahap langkah penelitian sebagai berikut:

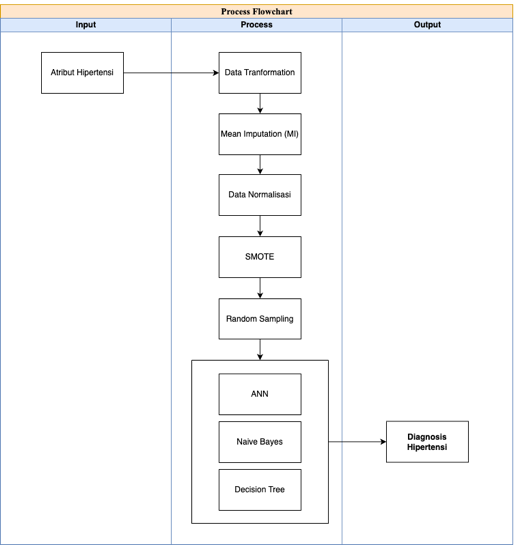


Figure 3 Process Flowchart

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.

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)**

Hasil yang diperoleh dengan menerapkan dataset hipertensi menggunakan algoritma Artificial Neural Network (ANN) serta menggunakan metode SMOTE dan tanpa SMOTE selanjutnya akan diuji dengan menggunakan random sampling yang berbeda yaitu 90:10, 80:20, dan 70:30.

Table 2 Hasil Evaluasi Menggunakan Metode ANN dalam persen

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RS** | **Dengan SMOTE** | | | | **Tanpa SMOTE** | | | |
|  | **Acc** | **Prec** | **Rec** | **F1-S** | **Acc** | **Prec** | **Rec** | **F1-S** |
| **9:1** | 77.7 | 14.7 | 77.7 | 82.8 | 93.4 | 100 | 93.4 | 90.3 |
| **8:2** | 76.7 | 16.4 | 76.7 | 82.7 | 94.1 | 100 | 94.1 | 91.2 |
| **7:3** | 74.3 | 15.6 | 74.6 | 81.1 | 93.8 | 100 | 93.8 | 90.9 |
| **Avg** | 76.3 | 15.5 | 76.3 | 82 | **93.7** | 100 | 93.7 | 90.8 |
| **RT** | 2.75(s) | | | | 1.06(s) | | | |

Acc : Accuracy

Prec : Precission

Rec : Recall

F1-S : F1-Score

Avg : Average

RT : Running Time

RS : Random Sampling.

Tabel 2 menunjukkan bahwa akurasi rata-rata adalah 76,3% dengan waktu komputasi 2,75 detik Ketika metode SMOTE diterapkan, sedangkan tanpa SMOTE akurasi rata-rata adalah 93,7% dengan waktu komputasi 1.06 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 dengan menerapkan dataset hipertensi menggunakan algoritma Naïve Bayes serta menggunakan metode SMOTE dan tanpa SMOTE selanjutnya akan diuji dengan menggunakan random sampling yang berbeda yaitu 90:10, 80:20, dan 70:30.

Table 3 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** | 62.0 | 8.62 | 62.0 | 71.6 | 88.8 | 26.6 | 88.8 | 90.3 |
| **8:2** | 60.4 | 8.8 | 60.4 | 70.7 | 90.1 | 26.9 | 90.1 | 91.2 |
| **7:3** | 63.5 | 12.0 | 63.5 | 72.9 | 90.3 | 26.4 | 90.8 | 90.9 |
| **Avg** | 61.9 | 9.80 | 61.9 | 71.7 | **89.7** | 26.6 | 90.5 | 90.8 |
| **RT** | 0.004(s) | | | | 0.001(s) | | | |

Tabel 3 menunjukkan bahwa akurasi rata-rata adalah 61,9% dengan waktu komputasi 0,004 detik Ketika metode SMOTE diterapkan, sedangkan tanpa SMOTE akurasi rata-rata adalah 89,7% dengan waktu komputasi 0.001 detik.

Kelebihan Naïve Bayes dapat digunakan untuk data kuantitatif dan kualitatif. Tidak perlu melakukan banyak pelatihan data. Jika ada nilai yang hilang, maka dapat diabaikan dalam perhitungan, dan perhitungannya cepat dan efisien sehingga mudah dimengerti. Sedangkan kelemahan dari Naïve Bayes adalah mengansumsikan bahwa setiap variabel bersifat independent sehingga mengurangi akurasi karena biasanya terdapat korelasi antara satu variabel dengan variabel lainnya.

**3.5. Decision Tree**

Hasil yang diperoleh dengan menerapkan dataset hipertensi menggunakan algoritma Decision Tree serta menggunakan metode SMOTE dan tanpa SMOTE selanjutnya akan diuji dengan menggunakan random sampling yang berbeda yaitu 90:10, 80:20, dan 70:30.

Table 4 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** | 88.8 | 23.0 | 88.8 | 89.5 | 88.8 | 18.1 | 88.8 | 93.1 |
| **8:2** | 88.2 | 17.8 | 88.2 | 89.4 | 92.4 | 33.3 | 92.4 | 92.1 |
| **7:3** | 89.5 | 22.2 | 89.5 | 90.1 | 91.7 | 30.7 | 91.7 | 91.5 |
| **Avg** | 88.8 | 21 | 88.8 | 89.6 | **90.9** | 27.3 | 90.9 | 92.2 |
| **RT** | 0.009(s) | | | | 0.007(s) | | | |

Tabel 3 menunjukkan bahwa akurasi rata-rata adalah 88,8% dengan waktu komputasi 0,009 detik Ketika metode SMOTE diterapkan, sedangkan tanpa SMOTE akurasi rata-rata adalah 90,9% dengan waktu komputasi 0.007 detik.

Kelebihan Decision Tree adalah sifatnya yang fleksibel sehingga mampu meningkatkan kualitas keputusan yang dihasilkan. Sedangkan kelemahan dari Decision Tree adalah akan terjadi overlap jika menggunakan data yang memiliki kelas dan kriteria dengan jumlah yang sangat banyak.

# Conclusion

Berdasarkan analisis dan pembahasan yang telah dilakukan terhadap data training dan testing menggunakan random sampling 90:10, 80:20, 70:30 dimana terdapat 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), Diagnosa 1, maka diperoleh kesimpulan bahwa:

1. hasil pengujian yang telah dilakukan, penggunaan SMOTE tidak memberikan performa dan peningkatan akurasi yang signifikan jika dibandingkan dengan penerapan dataset tanpa SMOTE pada keseluruhan pengujian, hal ini terlihat dari waktu komputasi yang diberikan. Dengan penggunaan SMOTE, waktu eksekusi yang dibutuhkan menjadi lebih banyak dikarenakan adanya penambahan data sintetis untuk menyeimbangkan kelas data minoritas.
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 93.7% dibandingkan dengan Naïve Bayes 89.7%, dan Decision Tree 90.9% 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.