

PENERAPAN DATA MINING DALAM MENGLASIFIKASI RESIKO PENYAKIT STROKE MENGGUNAKAN ALGORITMA SVM DAN ANN

Sitti Nurfatimah Bakari¹, Alter Lasarudin^{2*}, Wahyudin Hasyim³

^{1,2,3}Sistem Informasi Universitas Muhammadiyah Gorontalo

*Corresponding: sittybakari273@email.com**

Submitted: 09-05-2025; Accepted: 04-07-2025; Published: 12-07-2025

ABSTRACT

Stroke is a neurological condition characterized by the rapid onset of clinical signs, including severe local and global neurological deficits that persist for 24 hours or more and can be fatal when no clear cause other than vascular is identified. This study aimed to implement the *Support Vector Machine* (SVM) and *Artificial Neural Network* (ANN) algorithms to classify stroke risk and measure the accuracy of each algorithm. The study using the *Support Vector Machine* and *Artificial Neural Network* algorithms with a dataset of 200 records yielded an accuracy value of 97.46% for the *Support Vector Machine* and 98.75% for the *Artificial Neural Network*. The findings indicated that the *Artificial Neural Network* algorithm achieved the highest accuracy compared to the *Support Vector Machine* algorithm in classifying stroke risk data.

Keywords: Classification, Stroke Risk Based on Risk Factors, Support Vector Machine, Artificial Neural Network

ABSTRAK

Stroke merupakan kondisi neurologis yang ditandai dengan tanda klinis yang berkembang pesat, termasuk gangguan neurologis lokal dan global yang parah, bertahan selama 24 jam atau lebih, dan berakibat fatal ketika tidak ada penyebab yang jelas selain vaskular. Penelitian ini bertujuan untuk menerapkan algoritma *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) untuk mengklasifikasi resiko penyakit stroke dan mengukur tingkat akurasi dari masing-masing algoritma. Hasil penelitian menggunakan algoritma *Support Vector Machine* dan *Artificial Neural Network* dengan jumlah record data 200 maka akan diperoleh nilai akurasi *Support Vector Machine* sebesar 97.46%, dan tingkat akurasi *Artificial Neural Network* sebesar 98.75%. Terlihat algoritma *Artificial Neural Network* mendapatkan nilai akurasi yang tertinggi dibandingkan algoritma *Support Vector Machine* dalam pengklasifikasian data resiko penyakit stroke.

Kata Kunci: Klasifikasi, Resiko Stroke Berdasarkan Faktor Resiko, Support Vector Machine, Artificial Neural Network.

1. PENDAHULUAN

Stroke adalah kondisi neurologis yang ditandai dengan tanda klinis yang berkembang pesat, termasuk gangguan neurologis lokal dan global yang dapat parah, bertahan selama 24 jam atau lebih, dan/atau berakibat fatal ketika tidak ada penyebab yang jelas selain vaskular. Ketika arteri darah di otak pecah atau tersumbat, sebagian dari otak tidak dapat menerima aliran darah yang membawa oksigen, menyebabkan kematian jaringan dan sel, ini dikenal sebagai stroke [1] Organisasi Kesehatan Dunia melaporkan bahwa kelemahan wajah yang tiba-tiba dan bahkan mati rasa di lengan, kaki, atau di satu sisi tubuh adalah gejala yang paling sering dialami. Selain itu, fungsi sensorik tubuh hilang, dan sakit kepala yang sangat parah dirasakan, yang dapat menyebabkan pingsan atau tidak sadarkan diri [2].

Penyakit stroke di Gorontalo mengalami peningkatan, dari 7% menjadi 10,9% pada tahun 2018. Menurut data Rekam Medik RSUD Dr. M.M Dunda Limboto Kabupaten Gorontalo Kabupaten Gorontalo, diketahui bahwa 5 tahun terakhir yakni sejak tahun 2018 - 2023 ada kurang lebih 500 pasien penyakit stroke. Dan data yang didapat ada sekitar 200 pasien yang di rawat di RSUD Dr. M.M Dunda Limboto Kabupaten Gorontalo. Jika dilihat dari data yang ada faktor usia penderita stroke di RSUD Dr. M.M Dunda Limboto Kabupaten Gorontalo Kabupaten Gorontalo rata-rata berusia 23 tahun ke atas.

Alternatif solusi yang dilakukan pada penelitian ini adalah menerapkan metode *Machine Learning* dalam memanfaatkan data penyakit stroke untuk digali informasi atau pengetahuan baru didalamnya, sehingga dapat dimanfaatkan untuk membangun model baru dalam mengklasifikasi pasien untuk mengetahui apakah pasien tersebut menderita penyakit stroke yang beresiko Tinggi, Sedang, atau Rendah. *Machine learning* berfokus pada pengembangan sistem yang mampu belajar secara otomatis dan membuat keputusan tanpa memerlukan pemrograman ulang oleh manusia[3]. Proses penggalian informasi baru dari sekumpulan data dikenal dengan data mining. Algoritma *Machine Learning* yang digunakan untuk mengklasifikasi yaitu *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN). *Data mining* menjadi metode utama untuk menganalisis pola dan pengetahuan dari data medis guna meningkatkan ketepatan dan efisiensi dalam pengambilan keputusan klinis. Sejumlah algoritma telah diterapkan dalam konteks ini, termasuk *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN), yang dikenal memiliki tingkat akurasi tinggi dalam proses klasifikasi. *Data mining*, yang juga dikenal sebagai *Knowledge Discovery in Database* (KDD) merupakan serangkaian proses yang mencakup pengumpulan data, pembersihan, transformasi atau penyesuaian data, eksplorasi atau penggalian informasi, hingga penyajian hasil dalam kumpulan data berukuran besar [4]

Penelitian ini bertujuan untuk menerapkan metode *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) dalam mengklasifikasi resiko penyakit stroke. Dengan menggunakan Teknik *machine learning* dan data yang relevan, diharapkan dapat dikembangkan model klasifikasi yang akurat dan handal. Algoritma *Support Vector Machine* (SVM) sangat efektif dalam menganalisis data dan mentransformasi data ke dalam ruang fitur yang lebih besar, di mana pemisahan kelas menjadi lebih mudah dengan menggunakan kernel yang sesuai[5]. Sementara itu, *Artificial Neural Network* (ANN) sebuah sistem analisis data yang terinspirasi dari konfigurasi otak manusia [6], memungkinkan algoritma ini mengidentifikasi keterkaitan kompleks antar atribut dalam sebuah dataset, sehingga sangat ideal untuk pengolahan data yang kompleks dan bervariasi. Dengan menggunakan aplikasi *RapidMiner* sebagai solusi untuk menganalisis penambahan data, penambahan teks, dan analitik prediktif. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediktif untuk memberikan wawasan kepada pengguna dan membantu mereka membuat keputusan terbaik [7].

Pada penelitian sebelumnya dengan menggunakan sampel data alumni dari beberapa Universitas dan Institut yang ada di Kota Palembang. Hasil data uji yang didapat memiliki tingkat Akurasi sebesar 0.8506 atau sebesar 85.06% dengan pembagian data 50% data training dan 50% data testing yang digunakan dari 1739 record data dengan menggunakan model *Support Vector Machine* (SVM) menghasilkan akurasi sebesar 85.06% [8]. Dan peneliti [9] menyimpulkan bahwa Model jaringan ANN yang dapat dan baik digunakan dalam melakukan klasifikasi curah hujan di Jawa Tengah menunjukkan bahwa ketiga model tersebut memiliki nilai kesalahan yang sama dalam melakukan klasifikasi curah hujan yaitu sebesar 0.198. Berdasarkan peneliti [10] untuk mengklasifikasikan Indeks Pembangunan Manusia (IPM) di Pulau Sumatera, Berdasarkan hasil analisis pada penelitian ini didapatkan nilai akurasi untuk metode *Artificial Neural Network* (ANN) sebesar 97.4%, sedangkan nilai akurasi untuk metode *Support Vector Machine* (SVM) sebesar 53.25%. Sehingga SVM termasuk algoritma yang baik untuk digunakan [11]. Dari kedua nilai akurasi pada masing-masing metode tersebut maka dapat ditarik kesimpulan bahwa metode *Artificial Neural Network* (ANN) dapat digunakan untuk mengklasifikasikan Indeks Pembangunan Manusia (IPM) di Pulau Sumatera karena nilai akurasi pada metode *Artificial Neural Network* (ANN) lebih tinggi dibandingkan nilai akurasi pada metode *Support Vector Machine* (SVM).

2. METODE PENELITIAN

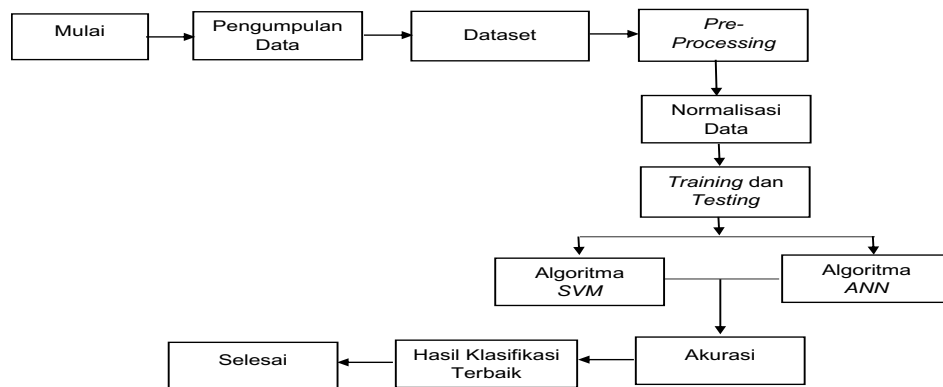
Objek Penelitian

Algoritma *Support Vector Machine* dan *Artificial Neural Network* seperti yang diketahui tidak berdiri sendiri. Penelitian ini dimulai dengan mengumpulkan data set karena tujuan penelitian ini untuk mengetahui seberapa akurat penggunaan algoritma *Support Vector Machine* dan *Artificial Neural Network*. Data set ini sesuai dengan rencana awal menggunakan data yang diambil langsung dari Rumah Sakit M.M Dunda Limboto.

Metode Penelitian

Penelitian ini dilakukan melalui beberapa langkah yaitu *observasi data set*, *preprocessing*, *training data*, dan *testing data*, *akurasi*, dan penjabaran hasil yang di dapatkan. Pengolahan data ini dilakukan dengan metode algoritma *Support Vector Machine* dan *Artificial Neural Network* untuk pengujian data set menggunakan tool *Rapid Miner*.

Tahapan Penelitian



Gambar 1 Tahapan Penelitian

Tahapan Perhitungan Algoritma SVM Linear

Algoritma *Support Vector Machine (SVM) Linear* untuk klasifikasi adalah metode pembelajaran mesin yang mencari garis atau bidang (*hyperplane*) terbaik untuk memisahkan data ke dalam dua kelas secara linier dengan memaksimalkan margin antara data dari kedua kelas tersebut.

Adapun data yang dikumpulkan sebagai contoh adalah sebagai berikut :

Ada 9 atribut dari dataset yaitu Jenis kelamin, Umur, Hipertensi, Diabetes mellitus, Hiperkolesterlemia, Penyakit jantung, Merokok, Obesitas, dan Asam Urat. Jadi formulasi yang digunakan adalah sebagai berikut :

$$\frac{1}{2} \| w \|^2 = \frac{1}{2} (w_1^2 + w_2^2 + w_3^2 \dots w_9^2) \quad \dots\dots (1)$$

Dengan syarat :

$$y_i (w \cdot X_i + b) \geq 1 \quad i=1,2,3,4,5 \dots n \dots\dots\dots (2)$$

$$y_i (w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + w_4 \cdot x_4 + w_5 \cdot x_5 + w_6 \cdot x_6 + w_7 \cdot x_7 + w_8 \cdot x_8 + w_9 \cdot x_9) \geq +1$$

Tinggi..... (3)

$$y_i (w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + w_4 \cdot x_4 + w_5 \cdot x_5 + w_6 \cdot x_6 + w_7 \cdot x_7 + w_8 \cdot x_8 + w_9 \cdot x_9) \geq -1$$

Rendah..... (4)

Tahapan Perhitungan Algoritma ANN Backpropagation

Artificial Neural Network Backpropagation Adalah Metode yang digunakan untuk melatih jaringan syaraf tiruan dengan cara mengoptimalkan bobot melalui proses perhitungan gradien yang terdiri dari Forward pass dan Backward pass.

Rumus persamaan yang digunakan pada *Artificial Neural Network Backpropagation* sebagai berikut :
Penjumlahan bobot menggunakan persamaan sebagai berikut:

$$z\_in_j = \sum_{i=1}^n X_i V_{ij} \dots\dots\dots (5)$$

Pengaktifan bobot menggunakan persamaan:

$$Z_j = \frac{1}{1+e^{-z\_in_j}} \dots\dots\dots (6)$$

Operasi pada output layer:

Perkalian bobot menggunakan persamaan:

$$y\_in_k = w_{ok} \sum_{i=1}^n Z_i W_{jk} \dots\dots\dots (7)$$

Pengaktifan bobot menggunakan persamaan:

$$y_k = \frac{1}{1+e^{-y\_in_k}} \dots\dots\dots (8)$$

Tahap Backpropagation:

Hitung faktor kesalahan dengan rumus:

$$\delta_k = (T_k - y_k) * (y_k) * (1 - y_k) \dots\dots\dots (9)$$

Koreksi bobot pada unit K:

$$\Delta w_{jk} = \alpha \delta_k Z_j \dots\dots\dots (10)$$

Perbaharui bobot dan prasikap:

$$\delta\_in_k = \sum_{k=1}^m \delta_k W_{jk} \dots\dots\dots (11)$$

$$\delta_k = \delta_{in} * \left(\frac{1}{1+e^{-z_j}} \right) * \left(1 - \left(\frac{1}{1+e^{-z_j}} \right) \right) \dots\dots\dots (12)$$

Perbaharui bobot tiap unit keluaran:

$$\Delta V_{jk} = \alpha \delta_k X_1 \dots\dots\dots (13)$$

Perubahan bobot input menuju hidden:

$$\Delta V_{j0} = \alpha \delta_k \dots\dots\dots (14)$$

Perbaharui bobot awal input ke hidden:

$$V_{ij}(\text{Baru}) = V_{ij}(\text{lama}) + \Delta V_{ij} \dots\dots\dots (15)$$

Perbaharui bobot awal hidden ke ouput:

$$W_{jk}(\text{Baru}) = W_{jk}(\text{lama}) + \Delta w_{jk} \dots\dots\dots (16)$$

$$V_{j0}(\text{Baru}) = V_{j0}(\text{lama}) + \Delta w_{j0} \dots\dots\dots (17)$$

$$W_{j0}(\text{bias baru}) = W_0(\text{lama}) + \Delta w_{j0} \dots\dots\dots (18)$$

3. HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang digunakan dalam penelitian adalah data yang diambil langsung dari RS M.M Dunda Limboto yang melalui observasi dan wawancara. Data yang diperoleh mencakup data pasien penyakit stroke berdasarkan faktor

Tabel 1 Dataset Awal

resiko stroke selama 5 tahun, dari tahun 2018 sampai 2023. Data ini di olah sehingga mendapatkan hasil klasifikasi dari penyakit stroke. Dataset ini merupakan tabel yang memuat informasi tentang jumlah pasien resiko penyakit stroke berdasarkan faktor resiko stroke.

No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterleミア (> 200)	Penyakit Jantung	Merokok	Obesitas (> 30)	Asam Urat (L : > 7, P : > 6)	Resiko Stroke
1	Laki-Laki	57	Ya	Tidak	272 mg/dl	Ya	Ya	28.4	10.7 mg/dl	Tinggi
2	Laki-Laki	61	Ya	Tidak	169 mg/dl	Tidak	Ya	21.5	9.0 mg/dl	Rendah
3	Laki-Laki	60	Tidak	Ya	230 mg/dl	Tidak	Ya	22.8	5.3 mg/dl	Rendah
4	Perempuan	39	Tidak	Ya	126 mg/dl	Tidak	Tidak	22.1	3.1 mg/dl	Rendah
5	Perempuan	64	Ya	Tidak	253 mg/dl	Tidak	Ya	25	4.5 mg/dl	Sedang
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
196	Perempuan	71	Ya	Ya	196 mg/dl	Tidak	Ya	30.2	6.1 mg/dl	Sedang
197	Perempuan	50	Tidak	Tidak	175 mg/dl	Tidak	Tidak	30.9	4.1 mg/dl	Rendah
198	Perempuan	68	Tidak	Ya	185 mg/dl	Tidak	Tidak	28.9	4.6 mg/dl	Rendah
199	Perempuan	56	Tidak	Ya	154 mg/dl	Tidak	Tidak	30.2	6.1 mg/dl	Rendah
200	Laki-Laki	79	Ya	Ya	266 mg/dl	Tidak	Ya	28.3	10.2 mg/dl	Tinggi

Sumber Data: Rumah Sakit M.M Dunda Limboto Tahun 2024

Preprocessing

Pre-processing bertujuan untuk mengolah data awal menjadi data yang lebih sederhana atau mengubah data yang berupa kategori menjadi numerik agar lebih mudah untuk diolah. Adapun atribut-atribut yang diubah menjadi numerik yaitu **Jeniskelamin**, **Hipertensi**, **Diabetes mellitus**, **Penyakit jantung**, dan **merokok**. Pada tahap ini atribut-atribut yang ada pada dataset ini berupa kategori Ya, diubah menjadi 2 dan kategori Tidak diubah menjadi 1.

Tabel 2 Data Preprocessing

No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterleミア (> 200)	Penyakit Jantung	Merokok	Obesitas (> 30)	Asam Urat (L : > 7, P : > 6)	Resiko Stroke
1	1	57	2	1	272	2	2	28.4	10.7	Tinggi
2	1	61	2	1	169	1	2	21.5	9	Rendah
3	1	60	1	2	230	1	2	22.8	5.3	Rendah
4	2	39	1	2	126	1	1	22.1	3.1	Rendah
5	2	64	2	1	253	1	2	25	4.5	Sedang
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
196	2	71	2	2	196	1	2	30.2	6.1	Sedang
197	2	50	1	1	175	1	1	30.9	4.1	Rendah
198	2	68	1	2	185	1	1	28.9	4.6	Rendah
199	2	56	2	1	154	1	1	30.2	6.1	Rendah
200	1	79	2	2	266	1	2	28.3	10.2	Tinggi

Setelah dilakukan *pre-processing*, data diubah yang awalnya label memiliki 3 kelas (tinggi, sedang, rendah) diubah menjadi 2 kelas yaitu tinggi dan rendah, hal ini dilakukan untuk melancarkan pengelolaan hasil klasiikasi, agar pengelolaan data dengan algoritma *Support vector Machine* berjalan dengan baik. Pembagian kelas tersebut ditemukan dengan cara *Median* menggunakan rumus :

$$x = \text{nilai data}$$

$$n = \text{jumlah data}$$

$$= \frac{x}{n} = \text{hasil median}$$

Tabel 3 Transformasi Data

No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterleミア (> 200)	Penyakit Jantung	Merokok	Obesitas (> 30)	Asam Urat (L : > 7, P : > 6)	Resiko Stroke
1	1	57	2	1	272	2	2	28.4	10.7	Tinggi
2	1	61	2	1	169	1	2	21.5	9	Rendah
3	1	60	1	2	230	1	2	22.8	5.3	Tinggi
4	2	39	1	2	126	1	1	22.1	3.1	Rendah
5	2	64	2	1	253	1	2	25	4.5	Tinggi
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
196	2	71	2	2	196	1	2	30.2	6.1	Tinggi
197	2	50	1	1	175	1	1	30.9	4.1	Rendah
198	2	68	1	2	185	1	1	28.9	4.6	Tinggi
199	2	56	2	1	154	1	1	30.2	6.1	Rendah
200	1	79	2	2	266	1	2	28.3	10.2	Tinggi

Normalisasi Data

Normalisasi data merupakan kegiatan mengubah nilai-nilai dalam dataset agar berada dalam rentang yang sama, biasanya 0 sampai 1.

Berikut rumus untuk melakukan normalisasi data yang digunakan :

$$X_i \frac{0,8 (xi - \min)}{(\max - \min)} + 0,1 \dots \dots \dots (3.1)$$

X_i = Nilai data

X' = Nilai hasil normalisasi

Min(x) = Nilai minimal dari seluruh atribut

Max(x) = Nilai maximal dari seluruh atribut

Tabel 4 Normalisasi Data

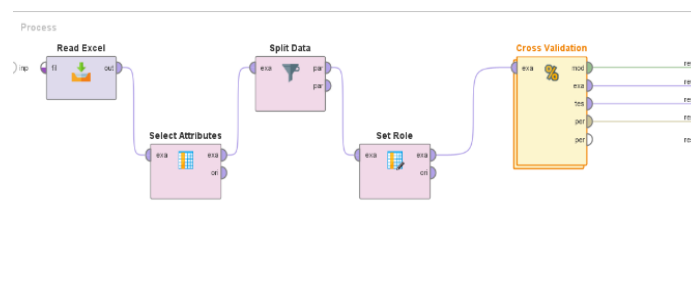
No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterolemia (> 200)	Penyakit Jantung	Merokok	Obesitas (> 30)	Asam Urat (L : > 7, P : > 6)	Resiko Stroke
1	0.100	0.561	0.900	0.100	0.622	0.900	0.900	0.687	0.891	Tinggi
2	0.100	0.615	0.900	0.100	0.273	0.100	0.900	0.329	0.731	Rendah
3	0.100	0.602	0.100	0.900	0.480	0.100	0.900	0.396	0.382	Tinggi
4	0.900	0.317	0.100	0.900	0.127	0.100	0.100	0.360	0.175	Rendah
5	0.900	0.656	0.900	0.100	0.558	0.100	0.900	0.510	0.307	Tinggi
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
196	0.900	0.751	0.900	0.900	0.364	0.100	0.900	0.781	0.458	Tinggi
197	0.900	0.466	0.100	0.100	0.293	0.100	0.100	0.817	0.269	Rendah
198	0.900	0.710	0.100	0.900	0.327	0.100	0.100	0.713	0.316	Tinggi
199	0.900	0.547	0.900	0.100	0.222	0.100	0.100	0.781	0.458	Rendah
200	0.100	0.859	0.900	0.900	0.602	0.100	0.900	0.682	0.844	Tinggi

Training dan Testing

Sebelum memulai pelatihan (*training*), perlu menentukan nilai parameter, jumlah data training, kelas data pelatihan (*training*), dan data pengujian (*testing*) sebagai langkah awal penyimpanan atribut vektor-vektor dan kelas klasifikasi data. Tujuan dari proses *testing* adalah untuk mendapatkan *output* kelas yang ditargetkan, yang kemudian dibandingkan dengan kelas klasifikasi pada data pengujian untuk menentukan presentasi keakurasiannya. Pada penelitian ini pengujian algoritma *Support Vector Machine* dan *Artificial Neural Network*, pada penelitian ini saya menggunakan 80% data untuk Training dengan jumlah data 160 dan 20% data untuk Testing dengan jumlah data 40 dari 200 dataset.

Pengujian Algoritma Support Vector Machine (SVM)

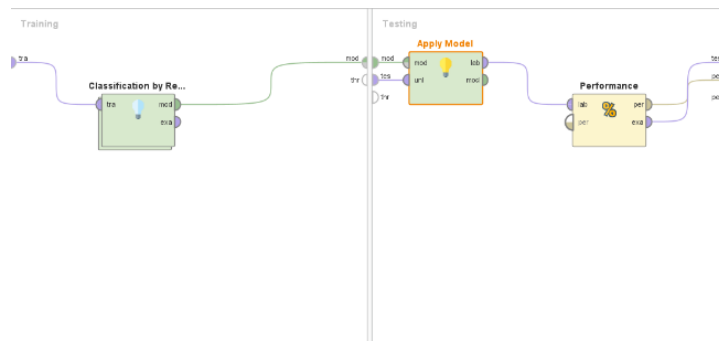
Pengujian software *RapidMiner* dilakukan untuk mengumpulkan informasi pengujian algoritma dari nilai akurasi yang ditujukan oleh *RapidMiner*.



Gambar 2 Pengujian Algoritma SVM Menggunakan RapidMiner

Berdasarkan gambar di atas menunjukkan bahwa pengelolaan dataset di rapidminer adalah sebagai berikut:

1. Memasukkan operator “Read Excel”, berfungsi memasukkan data yang akan diolah.
2. Memasukkan operator “Select Attributes” yang berfungsi menentukan atribut mana yang akan diklasifikasi.
3. Memasukkan Operator “Split Data”, berfungsi untuk membagi data *training* dan *testing*.
4. Memasukkan operator “Set Role”, digunakan untuk memilih target yang akan kita klasifikasi.
5. Menggunakan operator “Cross Validation”, untuk mengukur performa model secara lebih akurat dengan memanfaatkan seluruh dataset. Teknik ini digunakan untuk menghindari overfitting.

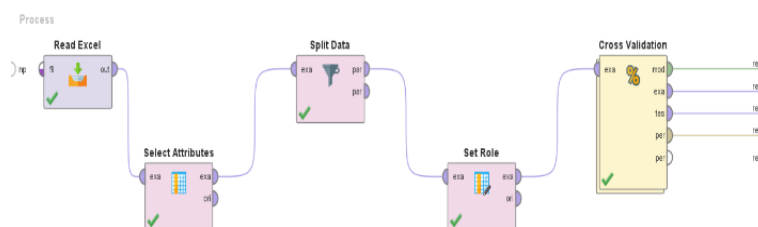


Gambar 3 Proses Operator Cross Validation

6. Memasukkan operator “Support Vector Machine Linear”, yang bertujuan untuk mengelola data, sehingga data tersebut mendapatkan hasil akurasi terbaik.
7. Memasukkan operator “Apply Model”, ini juga bertujuan untuk melihat hasil klasifikasi dari algoritma SVM
8. Memasukkan operator “Performance”, yang di gunakan untuk melihat hasil evaluasi seperti Akurasi.

Pengujian Algoritma Artificial Neural Network (ANN)

Selanjutnya pengujian algoritma *Artificial Neural Network* menggunakan software RapidMiner, operator-operator yang di gunakan dalam algoritma ini hampir sama dengan algoritma *Logistic Regression* hanya saja di dalam algoritma operator yang di gunakan pada bagian training testing diganti dengan operator *Neural Net* untuk pengujian algoritma *Artificial Neural Network* seperti pada gambar berikut:

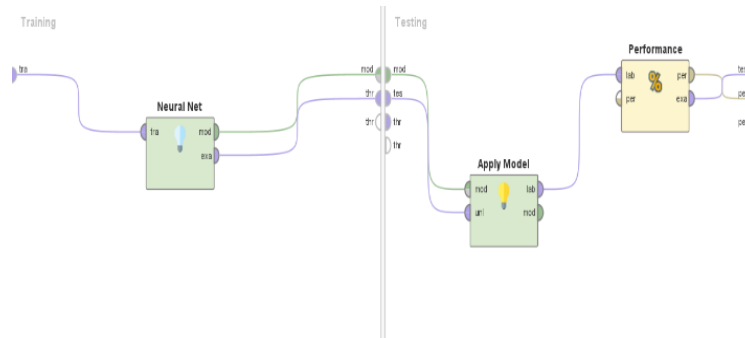


Gambar 4 Pengujian Algoritma ANN Menggunakan RapidMiner

Berdasarkan gambar di atas menunjukkan bahwa pengelolaan dataset di rapidminer adalah sebagai berikut:

1. Memasukkan operator “Read Excel”, berfungsi memasukkan data yang akan diolah.
2. Memasukkan operator “Select Attributes” yang berfungsi menentukan atribut mana yang akan diklasifikasi.
3. Memasukkan Operator “Split Data”, berfungsi untuk membagi data *training* dan *testing*.

4. Memasukkan operator “Set Role”, digunakan untuk memilih target yang akan kita klasifikasi.
5. Menggunakan operator “Cross Validation”, untuk mengukur performa model secara lebih akurat dengan memanfaatkan seluruh dataset. Teknik ini digunakan untuk menghindari overfitting.

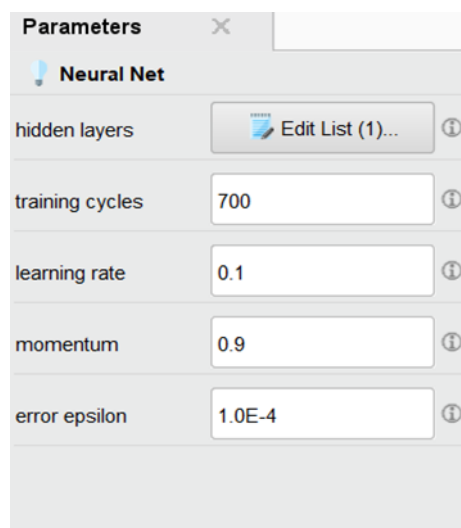


Gambar 5 Proses Operator Cross Validation

6. Memasukkan operator “Neural Net”, yang bertujuan untuk mengelola data, sehingga data tersebut mendapatkan hasil akurasi terbaik.
7. Memasukkan operator “Apply Model”, ini juga bertujuan untuk melihat hasil klasifikasi dari algoritma ANN
8. Memasukkan operator “Performance”, yang di gunakan untuk melihat hasil evaluasi seperti Akurasi.

Inisialisasi Parameter Algoritma Artificial Neural Network

Seluruh parameter dioptimalkan melalui metode percobaan guna memperoleh kinerja terbaik. Pemilihan parameter dilakukan dengan mempertimbangkan kemampuan model dalam menangani kompleksitas data sekuensial.



Gambar 6 Parameter Algoritma ANN

Pada gambar 6 menunjukkan parameter algoritma ANN yang harus dijalankan dalam aplikasi *RapidMiner* untuk mendapatkan nilai akurasi yang terbaik. Pada parameter yang pertama yaitu *Hidden Layer*, pada penelitian ini menggunakan skala 2 hingga 10, dan parameter yang kedua yaitu *Training Cycles*, menggunakan skala 100 hingga 1000, dan parameter yang ketiga yaitu *Learning Rate*, menggunakan skala 0.1 hingga 0.9 dan parameter yang keempat yaitu *Momentum*, menggunakan skala 0.1 hingga 0.9. Hasil optimasi menunjukkan bahwa parameter dengan hasil akurasi terbaik untuk algoritma ANN adalah **Hidden layer 5, Training Cycles 700,**

Learning Rate 0.1 , dan Momentum 0.9, Cross Validation 10 sehingga mendapatkan nilai akurasi terbaik yaitu **98.75%**.

Inisialisasi Parameter Algoritma Support Vector Machine

Seluruh parameter dioptimalkan melalui metode percobaan guna memperoleh kinerja terbaik. Pemilihan parameter dilakukan dengan mempertimbangkan kemampuan model dalam menangani kompleksitas data sekuensial.

Parameter	Value
C	0.01
convergence epsilon	0.001
L pos	0.1
L neg	100.0
epsilon	0.0
epsilon plus	0.0
epsilon minus	0.0

Gambar 7 Parameter Algoritma SVM

Pada gambar 7 menunjukkan parameter algoritma SVM yang harus dijalankan dalam aplikasi *RapidMiner* untuk mendapatkan nilai akurasi yang terbaik. Pada parameter yang pertama yaitu *C (Regularization parameter)*, pada penelitian ini menggunakan skala 0.0,0.01,0.1,1,10 , dan parameter yang kedua yaitu *Convergence Epsilon*, menggunakan skala 0.0001,0.001,0.1, dan parameter yang ketiga yaitu *L Pos*, menggunakan skala 0.01,0.1,1,10,100 dan parameter yang keempat yaitu *L Neg*, menggunakan skala 0.01,0.1,1,10,100. Hasil optimasi menunjukkan bahwa parameter dengan hasil akurasi terbaik untuk algoritma SVM adalah **C 0.01, Convergence Epsilon 0.001, Lpos 0.1, Lneg 1, Cross Validation 9**, sehingga mendapatkan nilai akurasi terbaik yaitu **97.46%**.

accuracy: 98.75% +/- 2.64% (micro average: 98.75%)

	true Tinggi	true Rendah	class precision
pred. Tinggi	80	0	100.00%
pred. Rendah	2	78	97.50%
class recall	97.56%	100.00%	

Gambar 8 Akurasi Algoritma Artificial Neural Network

accuracy: 97.46% +/- 5.05% (micro average: 97.50%)

	true Tinggi	true Rendah	class precision
pred. Tinggi	82	4	95.35%
pred. Rendah	0	74	100.00%
class recall	100.00%	94.87%	

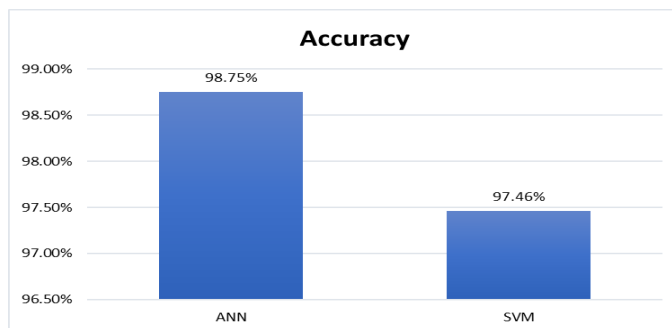
Gambar 9 Akurasi Algoritma Support Vector Machine

Pada gambar 8 dan 9 menjelaskan proses terakhir dalam pengujian mencari nilai akurasi tertinggi. Dengan melakukan pengujian menggunakan *Cross Validation* dan *Sampling type stratified sampling* dengan sepuluh kali pengujian, dan mengganti nilai *number of validation* 2 sampai 10 sehingga diperoleh nilai akurasi tertinggi.

Hasil Akurasi Algoritma SVM dan ANN

Dalam pengujian akurasi kedua algoritma, proses *training* menggunakan parameter yang sama. Hal ini terlihat pada pengujian menggunakan *Cross Validation sampling type stratified sampling* dengan sepuluh kali pengujian dan mengganti nilai *number of folds* 2 sampai 10 sehingga diperoleh nilai akurasi tertinggi. Hasil uji coba menunjukkan bahwa perbandingan dua algoritma ini didasarkan pada hasil klasifikasi dibandingkan dengan jumlah data *training* yang digunakan. Untuk mendapatkan hasil perbandingan dari kedua algoritma, nilai akurasi tertinggi rata – rata untuk setiap parameter dipilih.

Berdasarkan analisis data yang telah dilakukan maka didapatkan tingkat akurasi dari kedua algoritma dapat dilihat pada gambar dibawah ini :



Gambar 10 Perbandingan Akurasi Algoritma SVM dan ANN

Hasil uji coba dengan algoritma *Support Vector Machine* menunjukkan bahwa dari keseluruhan data, diperoleh hasil sehingga nilai akurasi sebesar 97.46%, dan uji coba dengan algoritma *Artificial Neural Network* menunjukkan bahwa dari keseluruhan data, diperoleh hasil sehingga nilai akurasi sebesar 98.75%. Jika ditinjau dari performance dalam melakukan proses running program, pada gambar 10 terlihat algoritma *Artificial Neural Network* mendapatkan nilai akurasi yang tertinggi dibandingkan algoritma *Support Vector Machine* dalam pengklasifikasian data resiko penyakit stroke.

Hasil Klasifikasi Algoritma Support Vector Machine (SVM)

Untuk hasil klasifikasi *Support Vector Machine* menggunakan data *testing* 20% atau 40 data. Dan menggunakan operator *Cross Validation* dengan *number of folds* 10 kali pengujian, jadi total data testing menjadi 400 data testing.

	true Tinggi	true Rendah	class precision
pred. Tinggi	82	4	95.35%
pred. Rendah	0	74	100.00%
class recall	100.00%	94.87%	

Gambar 11 Hasil Klasifikasi SVM

Berdasarkan gambar 11 untuk Resiko stroke Tinggi prediksi Tinggi ada 82 data. Dan Resiko stroke Rendah prediksi Rendah ada 74 data. Sedangkan Resiko stroke Tinggi prediksi Rendah ada 0 data. Dan Resiko stroke Rendah prediksi Tinggi ada 4 data.

Hasil Klasifikasi Algoritma Artificial Neural Network (ANN)

Untuk hasil klasifikasi *Artificial Neural Network* menggunakan data *testing* 20% atau 40 data. Dan menggunakan operator *Cross Validation* dengan *number of folds* 10 kali pengujian, jadi total data testing menjadi 400 data testing.

	true Tinggi	true Rendah	class precision
pred. Tinggi	80	0	100.00%
pred. Rendah	2	78	97.50%
class recall	97.56%	100.00%	

Gambar 12 Hasil Klasifikasi ANN

Berdasarkan gambar 12 untuk Resiko stroke Tinggi prediksi Tinggi ada 80 data. Dan Resiko stroke Rendah prediksi Rendah ada 78 data. Sedangkan Resiko stroke Tinggi prediksi Rendah ada 2 data. Dan Resiko stroke Rendah prediksi Tinggi ada 0 data.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan bahwa, Penerapan algoritma *Support Vector Machine (SVM)* dan *Artificial Neural Network (ANN)* dalam mengklasifikasi resiko penyakit stroke, kedua algoritma ini sangat baik digunakan karena algoritma SVM dan ANN mendapatkan hasil akurasi yang tinggi disesuaikan dengan data yang ada. Dengan melakukan pengujian hasil klasifikasi resiko penyakit stroke menggunakan 200 data, diperoleh nilai akurasi yang berbeda dari algoritma *Support Vector Machine* dan *Artificial Neural Network*. Tingkat akurasi algoritma *Support Vector Machine* sebesar 97.46% dan tingkat akurasi algoritma *Artificial Neural Network* sebesar 98.75%. Jika ditinjau dari performance dalam melakukan proses *running program*, terlihat algoritma *Artificial Neural Network* mendapatkan nilai akurasi yang tertinggi dibandingkan algoritma *Support Vector Machine* dalam pengklasifikasian data resiko penyakit stroke.

DAFTAR PUSTAKA

- [1] R. A. Saputra *et al.*, "Detecting Alzheimer's Disease by the Decision Tree Methods Based on Particle Swarm Optimization," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, pp. 61–67, 2020, doi: 10.1088/1742-6596/1641/1/012025.
- [2] Y. Azhar, A. K. Firdausy, and P. J. Amelia, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," vol. 5, no. 2, pp. 191–197, 2022.
- [3] R. Pratama, M. I. Herdiansyah, D. Syamsuar, and A. Syazili, "Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning," vol. 12, pp. 96–104, 2023.
- [4] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [5] A. Jalil, A. Homaidi, and Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2070–2079, 2024, doi: 10.33379/gtech.v8i3.4811.
- [6] S. Kasus, "Perbandingan Metode Support Vector Machine (SVM) dan Artificial Neural Network (ANN) pada Klasifikasi Gizi Balita," pp. 1036–1043, 2021.
- [7] N. R. Muntiari and K. H. Hanif, "Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning," vol. 3, no. 1, pp. 1–6, 2022.
- [8] M Riski Qisthiano, "Klasifikasi Terhadap Prediksi Kelulusan Mahasiswa Dengan Menggunakan Metode Support Vector Machine (Svm)," *Semin. Nas. Teknol. dan Multidisiplin Ilmu*, vol. 2, no. 2, pp. 203–207, 2022, doi: 10.51903/semnastekmu.v2i1.170.
- [9] N. K. Maulidia, "Penerapan artificial neural network pada proses klasifikasi curah hujan di Jawa Tengah," 2023, [Online]. Available: <http://etheses.uin-malang.ac.id/id/eprint/58140>
- [10] K. Indeks, P. Manusia, I. P. M. Di, M. Fathurrahman, and N. Qisthi, "Sumatera Pada Dataset Multi-Class Dengan Metode Artificial Neural Network (Ann)," vol. 0, pp. 377–384, 2021.

- [11] M. I. Abas and I. Ibrahim, "Optimasi Support Vector Machine Particle Swarm Optimization Untuk Prediksi Konsumsi Energi Listrik," *Jambura J. Informatics*, vol. 1, no. 2, pp. 47–56, 2019, doi: 10.37905/jji.v1i2.2646.