PENERAPAN DATA MINING DALAM MENGKLASIFIKASI RESIKO PENYAKIT STROKE MENGGUNAKAN METODE MACHINE LEARNING

Sitti Nurfatimah Bakari[1], Alter Lasarudin[2], Wahyudin Hasyim [3]*
[1,2,3]Program Studi Sistem Informasi, Universitas Muhammadiyah Gorontalo
sittybakari273@gmail.com[1*]
Universitas Muhammadiyah Gorontalo, Jl. Prof. Mansoer Pateda, Kab. Gorontalo, Indonesia

ABSTRACT

SITTI NURATIMAH BAKARI, Application of Data Mining in Classifying Stroke Risk at M.M Dunda Limboto Hospital Using Machine Learning Methods Supervised by Alter Lasarudin, as the first supervisor, and Wahyudin Hasyim , as the secondary supervisor.

Stroke is 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 Suppor Vector Machine and 98,75% for the Artficial 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

SITTI NURFATIMAH BAKARI, Penerapan Data Mining Dalam Mengklasifikasi Resiko Penyakit Stroke Menggunakan Metode Machine Learning. Dibimbing oleh Alter Lasarudin, M. Kom sebagai pembimbing 1 dan Wahyudin Hasyiim, M. Kom sebagai pembimbing 2.

Stroke adalah 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 *Artitiicial Neural Network* (ANN) untuk mengklasifikasi resiko penyakit stroke dan mengukur tingkat akurasi dari masing-masing algoritma. Hasil penelitian menggunakan algoritma *Support Vector Machine* dan *Artitiicial Neural Network* dengan jumlah record data 200 maka akan diperoleh nilai akurasi *Support Vector Machine* sebesar 97.46%, dan tingkat akurasi *Artitiicial 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.(Saputra et al., 2020).

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. 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).

Penelitian ini bertujuan untuk menerapkan metode machine learning dalam mengklasifikasi resiko penyakit stroke. Dengan menggunakan Teknik machine learning dan data yang relevan, diharapkan dapat dikembangkan model klaifikasi yang akurat dan handal.

2. TINJAUAN PUSTAKA

a. Machine Learning

Machine Learning adalah subbidang dari kecerdasan buatan (AI), menurut IBM. Ini adalah bidang ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk memecahkan masalah dan membuat keputusan, selain mensimulasikan bagaimana orang belajar dan secara bertahap meningkatkan akurasi belajar komputer. *Machine Learning* bertujuan untuk membuat sistem pembelajaran dan pengambilan keputusan secara mandiri yang tidak memerlukan pemrograman ulang manusia. Mesin dapat menyesuaikan diri dengan perubahan saat terjadi, selain menentukan tindakan terbaik saat membuat keputusan. (Pratama et al., 2023)

b. Data Mining

Data mining adalah proses ekstraksi dan identifikasi informasi yang relevan dan pengetahuan yang relevan dari kumpulan data besar menggunakan statistik, matematika, AI, dan machine learning. Selain mengumpulkan data, data mining juga melibatkan menganalisis dan meramalkan informasi yang akan ditampilkan. Data yang dikumpulkan pertama-tama diproses dan kemudian disimpan dalam database saat membuat keputusan tentang informasi yang akan digunakan. (Sains, 2022)

c. Klasifikasi

Klasifikasi adalah metode pengolahan data yang membagi data ke dalam kelompok berdasarkan kriteria tertentu. Dengan mengklasifikasikan data, kita dapat memastikan bahwa itu akurat. Dalam penambangan data, masalah klasifikasi dapat diidentifikasi dengan berbagai metode. (Akmal et al., 2023)

d. Support Vector Machine (SVM)

Menurut (Sang et al., 2021) *Support Vector Machine (SVM)*, sebuah sistem pembelajaran yang menggunakan ruang hipotesis dalam bentuk fungsi linier di ruang fitur berdimensi tinggi, dilatih menggunakan teknik pembelajaran yang didasarkan pada teori optimisasi.Parameter dan fungsi kernel memengaruhi ketepatan model yang dihasilkan oleh proses transisi *Support Vector Machine* . Pendekatan linier *Support Vector Machine* dan pendekatan *non-linier Support Vector Machine* membagi data yang dipisahkan secara *linier* (yaitu, membagi dua kelas pada *hyperplane* dengan margin lunak.

e. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) dapat digunakan untuk klasifikasi dan prediksi dalam Data mining. Dalam *Machine learning* tujuan untuk meniru *neurofisiologi* otak manusia dengan menggabungkan komponen komputasi dasar (neuron) ke dalam sistem jaringan. Dua area utama aktivitas algoritma *neural network* dan otak manusia adalah pengetahuan yang diperoleh melalui pembelajaran dan kekuatan koneksi antara neuron, atau bobot, yang berfungsi sebagai media untuk menyimpan informasi yang dipelajari.Jaringan saraf adalah algoritma yang sangat baik untuk memproses berbagai jenis data suara. (Purwono et al., 2022)

f. Stroke

Stroke adalah serangan mendadak pada otak yang mengakibatkan gangguan aliran darah yang disebabkan oleh penyumbatan atau pecahnya satu atau lebih pembuluh darah otak,yang baik sebagian atau seluruhnya mengganggu fungsi otak. Sel - sel otak mulai mengalami kekurangan darah, oksigen, atau nutrisi sebagai akibatnya, dan mereka akhirnya dapat mati dengan cepat. 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. (Azhar et al., 2022)

g. Penelitian Terkait

Penelitian yang dilakukan oleh (Indeks et al., 2021) dengan judul "Klasifikasi Indeks Pembangunan Manusia (IPM) di Pulau Sumatera Pada Dataset Multi-class Dengan Metode *Artificial Neural Network* (ANN)", menyimpulkan bahwa untuk membandingkan metode *Artificial Neural Network* dan *Support Vector Machine (SVM*) 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%. 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).*

3. METODE PENELITIAN

a. Objek Penelitian

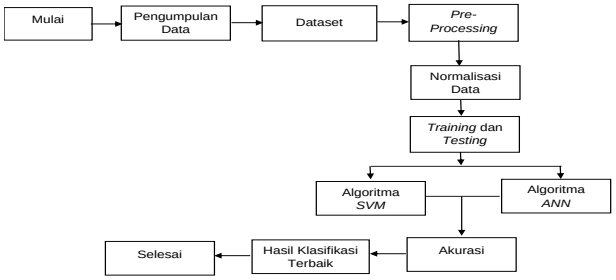
Algoritma *Support Vector Machine* dan *Artificial Neural Network* seperti yang diketahui tidak berdiri

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.

b. Metode Penelitian

Penelitian ini dilakukan melalui beberapa langkah yaitu o*bservasi data set, prepocessing, 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*.

c. Tahapan Penelitian



d. 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 :

$\frac{1}{2} \| w \|^2 = \frac{1}{2} (w_2^1 + w_2^2 + w_2^3 \ ... w_2^9)$ …… **(3.1)**

Dengan syarat :

$y_i \ (w . \ X_i \ + b \ ) \geq 1$ i=1,2,3,4,5…..n……**(3.2)**

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

Tinggi…… **(3.3)**

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

Rendah…… **(3.4)**

e. 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:

$z_in_j = \sum_{i=1}^{n} X_i \, V_{ij}$ …… **(3.5)**

Pengaktifan bobot menggunakan persamaan:

$Z_j = \frac{1}{1+e^{-z_inj}}$ …… **(3.6)**

Operasi pada output layer:

Perkalian bobot menggunakan persamaan:

$y_in_k = w_{ok} \sum_{i=1}^{n} Z_i \, W_{jk}$ …….**(3.7)**

Pengaktifan bobot menggunakan persamaan:

$y_k = \frac{1}{1+e^{-y_ink}}$ ……..**(3.8)**

Tahap Backpropagation:

Hitung faktor kesalahan dengan rumus:

$\delta_k = (T_k - y_k) * (y_k) * (1 - y_k)$ …..**(3.9)**

Koreksi bobot pada unit K:

$\Delta w_{jk=\alpha \delta_k} Z_j$ …….. **(3.10)**

Perbaharui bobot dan prasikap:

$\delta_in_k = \sum_{k=1}^{m} \delta_k \, W_{jk}$ …….. **(3.11)**

$\delta_k = \delta_{in} * \left(\frac{1}{1+ e^{-zj}}\right) * \left(1 - (\frac{1}{1+ e^{-zj}})\right)$ ….**(3.12)**

Perbaharui bobot tiap unit keluaran:

$\Delta V_{jk} = \alpha \delta_k X1$ …….. **(3.13)**

Perubahan bobot input menuju hidden:

$\Delta V_{j0} = \alpha \delta_k$ …….. **(3.14)**

Perbaharui bobot awal input ke hidden:

$V_{ij}(Baru) = V_{ij}(lama) + \ \Delta_{Vij}$ …**(3.15)**

Perbaharui bobot awal hidden ke ouput:

$W_{jk}(Baru) = W_{jk}(lama) + \ \Delta w_{jk}$…. **(3.16)**

$V_{j0}(Baru) = V_{j0}(lama) + \ \Delta w_{j0}$ …**(3.17)**

$W_{j0}(bias \ baru) = W_0(lama) + \ \Delta w_{j0}$ …**(3.18)**

4. HASIL DAN PEMBAHASAN

a. 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 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.

Tabel 1 Dataset Awal

No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterlemia (> 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	Ya	Tidak	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

b. 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 menjadi1.

Tabel 2 Data Preprocessing

No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterlemia (> 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 penggelolaan data dengan algoritma *Support vector Machine* berjalan dengan baik. Pembagian kelas tersebut ditemukan dengan cara *Median* menggunakan rumus :

x = nilai data

n = jumlah data

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

Tabel 3 Transformasi Data

No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterlemia (> 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

c. Normalisasi

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

Berikut rumus untuk melakukan normalisasi data yang digunakan :

$$X_i \frac{0{,}8\ (xi - \min)}{(\max - \min)} + 0{,}1 \ldots\ldots\ldots\ldots\ldots\ldots (4.1)$$

Xi = Nilai data

X' = Nilai hasil normalisasi

Min(x) = Nilai minimal dari seluruh atribut

Max(x) = Nilai maximal dari seluruh atribut

Tabel 4 Data Normalisasi

No.	Jenis Kelamin	Umur	Hipertensi	Diabetes Mellitus	Hiperkolesterlemia (> 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

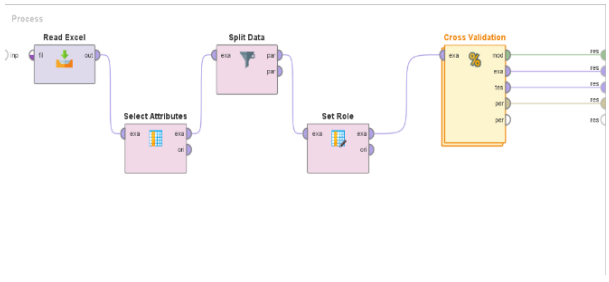
d. *Training* dan *Testing*

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 klasfikasi 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.

e. Pengujian Algoritma SVM dan ANN

1. Pengujian Algoritma *Support Vector Machine*

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

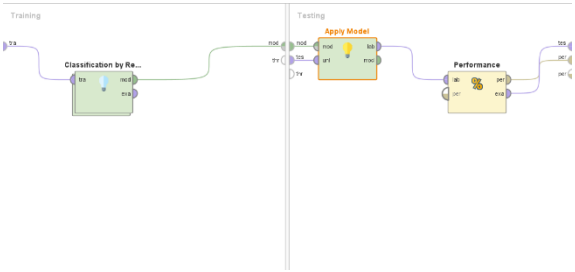


Gambar 1 Pengujian algoritma SVM menggunakan RapidMiner

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 menentukkan 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.

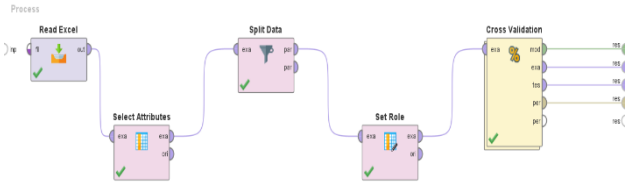


Gambar 2 Proses Operator Cross Validation

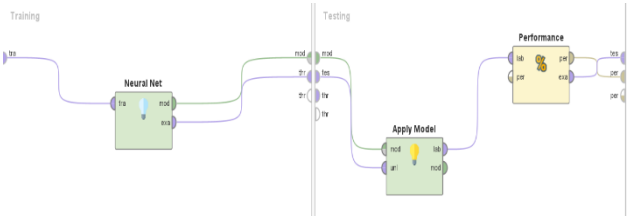
6. Memasukkan operator “*Support Vector Machine Linear”*, yang bertujuan untuk mengelola data, sehingga data tersebut mendapatkan hasil akurasi terbaik.
7. Memasukkan operator “Apply Model”, ini juga bertujuan untuk melihat hasil klasifikasi dari algoritma SVM .
8. Memasukkan operator “Performance”,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

2. Pengujian Algoritma *Artificial Neural Network*

Selanjutnya pengujian algoritma *Artificial Neural Network* menggunakan software RapidMiner, operator-operator yang di gunakan dalam algoritma ini hampir sama dengan algoritma *Logistic Regresion* 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 3 Pengujian algoritma ANN menggunakan RapidMiner



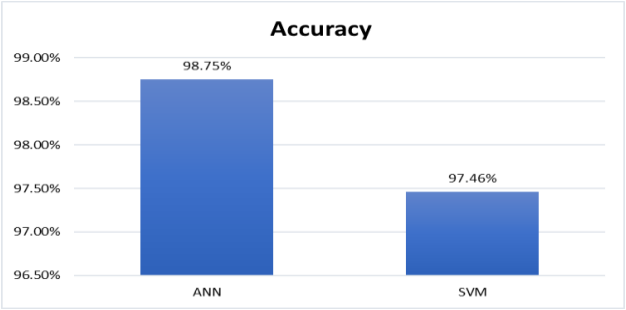
Gambar 4 Proses Operator Cross Validation

Seluruh parameter dioptimalkan melalui metode percobaan guna memperoleh kinerja terbaik. Pemilihan parameter dilakukan dengan mempertimbangkan kemampuan model dalam menangani kompleksitas data sekuensial. 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,** sedangkan untuk algoritma *ANN* adalah **Hidden layer 5, Training Cycles 700, Learning Rate 0.1 , dan Momentum 0.9, Cross Validation 10**.

f. 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.



Gambar 5 Hasil 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 5 terlihat algoritma *Artificial Neural Network* mendapatkan nilai akurasi yang tertinggi dibandingkan algoritma *Support Vector Machine* dalam pengklasifikasian data resiko penyakit stroke.

g. Hasil *Klasifikasi* Algoritma LR dan ANN

1. Hasil Klasifikasi Algoritma *Support Vector Machine*

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 6 Hasil Klasifikasi Algoritma SVM

Berdasarkan gambar 6 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.

2.* Hasil Klasifikasi Algoritma *Artificial Neural Network

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

Gambar 7 Hasil Klasifikasi Algoritma ANN

Berdasarkan gambar 7 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.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan :

1. Penerapan algoritma *Support Vector Machine (SVM)* dan *Artificial Neural Network (ANN)* dalam mengklasfikasi resiko penyakit stroke, kedua algoritma ini sangat baik digunakan karena algoritma SVM dan ANN mendapatkan hasil akurasi yang tinggi disesuaikan dengan data yang ada.
2. 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.

Dari hasil penelitian, maka terdapat saran yang dapat digunakan sebagai penelitian lebih lanjut terkait dengan penelitian ini adalah penelitian selanjutnya peneliti bisa menggunakan algoritma *Support Vector Machine* dan *Artificial Neural Network* untuk mengklasifikasi data menyesuaikan dengan data yang digunakan.

DAFTAR PUSTAKA

[1] Pratama, R., Herdiansyah, M. I., Syamsuar, D., & Syazili, A. (2023). *Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning*. *12*, 96–104.

[2] Sains, S. (2022). Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random. https://doi.org/10.33372/stn.v8i2.885

[3] Akmal, K., Faqih, A., Dikananda, F., Studi, P., Informatika, T., Studi, P., Informasi, S., & Neighbors, K. (2023). Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke. 7(1), 470–477.

[4] Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine Data Minning Analysis For Classification Of Air Quality Data Dki Jakarta Using Decision Tree Algorthm And Support Vector. 8(5), 8954–8963.

[5] Purwono, P., Dewi, P., Wibisono, S. K., Dewa, B. P., Informatika, P., Bangsa, U. H., Keperawatan, P., & Bangsa, U. H. (2022). Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network. 7(2), 82–90.

[6] Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. 5(2), 191–197.

[7] Indeks, K., Manusia, P., Di, I. P. M., Fathurrahman, M., & Qisthi, N. (2021). Sumatera Pada Dataset Multi-Class Dengan Metode Artificial Neural Network (ANN). 0, 377–384.

[8] Saputra, R. A., Agustina, C., Puspitasari, D., Ramanda, R., Warjiyono, Pribadi, D., Lisnawanty, & Indriani, K. (2020). Detecting Alzheimer's Disease by the Decision Tree Methods Based on Particle Swarm Optimization. Journal of Physics: Conference Series, 1641(1), 61–67. https://doi.org/10.1088/1742-6596/1641/1/0