



Implementation of Decision Tree Algorithm C4.5 and Support Vector Regression for Stroke Disease Prediction

Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke

**Firman Akbar^{1*}, Hanif Wira Saputra², Adhitya Karel Maulaya³,
Muhammad Fikri Hidayat⁴, Rahmaddeni⁵**

^{1,2,3,4,5}Program Studi Teknik Informatika, STMIK AMIK Riau, Indonesia

E-Mail: ¹2110031802067@sar.ac.id, ²2110031802063@sar.ac.id, ³2110031802059@sar.ac.id,
⁴2110031802062@sar.ac.id, ⁵rahmaddeni@sar.ac.id

Received Aug 20th 2022; Revised Sept 05th 2022; Accepted Sept 10th 2022
Corresponding Author: Rahmaddeni

Abstract

Data mining is the process of gathering important information and data from a large amount of data to be extracted and transforming it into new information useful for decision making. The data used in this article comes from neurological (neurological) patient data, specifically stroke, processed using support vector regression algorithms and C4.5 decision trees. Strokes are caused by ruptured blood vessels and blocked arteries in the brain, causing cells or tissue to die because they cannot supply the blood needed to carry oxygen to the brain. One way to study strokes is through data mining, which uses support vector regression and C4.5 decision tree algorithms. The results of this report identify stroke patients with known variables and analyze them using the C4.5 decision tree data mining algorithm and support vector regression. It can be seen that if the error produced by the C4.5 decision tree algorithm is 0.235 at the ratio of 70:30, then the support vector regression algorithm is 0.399 at the ratio of 70:30. Using the C4.5 decision tree algorithm, it will produce additional output in the form of a decision tree graph where there is a stream of predictions.

Keyword: Data Mining, Decision Tree C4.5, Prediction, Stroke, Super Vector Regression.

Abstrak

Data mining adalah proses pengumpulan informasi dan data penting dari sejumlah besar data yang perlu diekstraksi untuk mengubahnya menjadi informasi baru yang berguna untuk pengambilan keputusan. Data yang digunakan dalam penulisan ini berasal dari data pengidap neurologi (saraf) tepatnya stroke, diolah menggunakan algoritma *Support Vector Regression* dan *Decision Tree C4.5*. Stroke disebabkan oleh pecahnya pembuluh darah dan tersumbatnya pembuluh darah arteri di otak, sehingga mengakibatkan kematian sel atau jaringan karena tidak mensuplai darah yang dibutuhkan untuk membawa oksigen ke bagian otak. Suatu cara untuk meninjau stroke adalah data mining, yang memakai algoritma *Support Vector Regression* dan *Decision Tree C4.5*. Hasil laporan ini mengidentifikasi pengidap penyakit stroke pada variabel yang didapati dan menganalisisnya memakai algoritma data mining *Decision Tree C4.5* dan *Support Vector Regression*. Dapat dilihat jika error yang dihasilkan oleh algoritma *Decision Tree C4.5* terhadap rasio 70 : 30 bernilai 0.235, Selanjutnya untuk algoritma *Support Vector Regression* terhadap rasio 70 : 30 bernilai 0.399, Dalam menggunakan algoritma *Decision Tree C4.5*, maka akan menghasilkan output tambahan berupa sebuah grafik pohon keputusan dimana terdapat alur dalam memprediksi.

Kata Kunci: Data Mining, Decision Tree C4.5, Prediksi, Strok, Super Vector Regression.

1. PENDAHULUAN

Data mining sekarang menjadi salah satu titik fokus utama bagi para ilmuwan dan praktisi Data mining adalah proses pengumpulan informasi dan data penting dari sejumlah besar data yang perlu diekstraksi untuk mengubahnya menjadi informasi baru yang berguna untuk pengambilan keputusan [1]. Teknologi data mining tidak hanya digunakan dalam industri keuangan dan telekomunikasi, tetapi juga banyak digunakan dalam perawatan kesehatan. Misalnya, Mayo Clinic bekerja sama dengan IBM untuk menerapkan teknik penambangan data kepada pasien dengan keterangan jenis kelamin, usia, dan riwayat medis yang sama untuk

menentukan respons mereka terhadap suatu pengobatan tertentu. [2]. Dengan berkembangnya sistem informasi di bidang kesehatan membuat masyarakat dan rumah sakit bisa mendapatkan berbagai macam kemudahan. Penerapan sistem informasi di bidang medis berperan dalam menghindari malpraktik dalam penangkalan, respon dan umpan balik.

Dalam hal ini, studi kasus yang diselidiki adalah suatu penyakit berpotensi mematikan terhadap manusia, ialah stroke. Stroke ialah pemicu kematian nomor dua di dunia dan pemicu keterlatarbelakangan mental nomor tiga. Menurut Organisasi Kesehatan Dunia (*World Health Organization*), stroke adalah suatu kondisi yang ditandai dengan tanda-tanda klinis yang berkembang pesat berupa cacat neurologis lokal dan global yang dapat parah dan berlangsung selama 24 jam atau lebih dan atau dapat berakibat fatal, jika tidak ada penyebab jelas selain penyebab pembuluh darah. Stroke terjadi apabila pembuluh darah otak mengalami penyumbatan atau mengalami pecah yang menyebabkan sebagian otak tidak mendapatkan pasokan darah yang membawa oksigen yang dibutuhkan sehingga berakibat kematian sel/jaringan [3].

Hal ini mendorong banyak penelitian terhadap penyakit stroke, salah satunya memakai metode berbasis komputer. Metode ini dapat mengelola dataset yang besar untuk melakukan prediksi dengan algoritma tertentu sehingga hasilnya lebih cepat dan akurat. Prediksi adalah proses memperkirakan sesuatu secara sistematis berdasarkan informasi masa lalu dan saat ini yang tersedia, ini mungkin terjadi di masa depan sendiri, sehingga kesalahan (selisih antara yang terjadi dan yang diperkirakan) dapat diminimalisir [4]. Penelitian sebelumnya menggunakan algoritma Regresi Linear berganda dengan dataset yang sama dan menghasilkan nilai *Root Mean Squared Error* sebesar 0.218 +/- 0.000 [4]. Lalu algoritma *Decision Tree ID3* juga digunakan dalam memprediksi penyakit ISPA (Infeksi Saluran Pernapasan Akut) yang menghasilkan nilai *Error* sebesar 12.5% [5].

Penelitian ini merupakan penelitian lebih lanjut yang membandingkan algoritma *Support Vector Regression* dan *Decision Tree C4.5* terhadap data pasien stroke. Algoritma *Support Vector Regression* merupakan pengembangan dari algoritma *Support Vector Machine* untuk kasus regresi atau prediksi. Dibandingkan dengan model regresi lain seperti *Ordinary Least Squares*, *Support Vector Regression* memiliki kelebihan yaitu dapat menangani data yang *overfitting* sehingga algoritma ini memiliki kinerja yang bagus, sedangkan metode OLS memiliki kinerja yang buruk dalam hasil prediksi untuk data tersebut [6]. Algoritma *Decision Tree C4.5* merupakan salah satu algoritma yang populer dan mudah dipahami karena hasilnya seperti otak manusia sehingga mudah untuk memahami aturan yang didapat dari hasil tersebut. Algoritma ini bahkan berpengaruh di bidang medis, membantu dokter membuat keputusan penting untuk laporan patologi tertentu, karena sifat pohonnya, maknanya, dan kesimpulannya yang bahkan dapat dipahami oleh orang banyak [7].

Berdasarkan hal di atas, peneliti menggunakan 5110 dataset stroke yang diperoleh dari *Kaggle* untuk melakukan pengujian membandingkan algoritma *Support Vector Regression* dengan *Decision Tree C4.5*. Data ini kemudian diolah menggunakan *software RapidMiner* dan analisis data dilakukan secara otomatis. Penelitian ini akan digunakan sebagai acuan untuk memprediksi stroke. Hasil pengolahan data stroke diharapkan dapat memberikan informasi yang dapat menemukan potensi atau pengetahuan yang lebih baik dan menemukan peluang atau rencana strategis baru dalam prediksi stroke. Ini juga dapat digunakan sebagai sarana untuk membuat keputusan pencegahan stroke.

2. METODE PENELITIAN

Metode penelitian ini berlangsung dari berbagai tahap proses, mengumpulkan data ditujukan untuk penerapan dan menganalisis kadar yang ditujukan untuk mengolah data dengan membagikan data yang sudah diproses. Distribusi data latih dan data uji dengan perbandingan 70:30. Selanjutnya merancang model algoritma *Support Vector Regression* dan *Decision Tree C4.5*. Kemudian jalankan proses pelatihan data menggunakan algoritma yang akan gunakan. Pada tahap ini, model memeriksa fitur yang diperoleh dan mengubah nilai parameter selama pelatihan data. Langkah selanjutnya adalah mengeksplorasi model menggunakan data uji. Data pengujian kemudian memprediksi hasil dan membandingkannya dengan nilai parameter dari data pengujian yang sebenarnya untuk mendapatkan tingkat kesalahan dalam memprediksi proses pengujian data dan data pengujian yang sebenarnya. Adapun tahapan penelitian yang dilakukan pada penelitian ini ditunjukkan pada gambar 1.

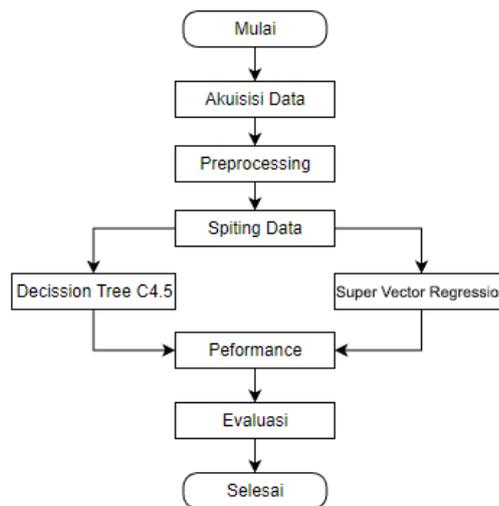
Perancangan model algoritma *Support Vector Regression* dan *Decision Tree C4.5* yang dibangun akan digunakan untuk memprediksi data pasien strok berdasarkan parameter yang diberikan.

2.1 Akuisisi Data

Penelitian adalah usaha yang meliputi tahapan dan dilakukan secara sistematis untuk mendapatkan jawaban ilmiah dari suatu masalah. Tentu saja dalam kasus penelitian komponen utamanya adalah kumpulan data, yang diproses menggunakan algoritma. Dataset yang digunakan dalam penelitian ini adalah dataset *Kaggle stroke*. Kumpulan data memiliki 12 fitur, yaitu: *id*, *gender*, *age*, *hypertension*, *heart_disease*, *ever_married*, *work_type*, *Residence_type*, *avg_glucose_level*, *bmi*, *smoking_status*, dan *stroke*. berikut tabel 1. yang berisikan deskripsi kumpulan data:

Tabel 1. Deskripsi Kumpulan Data

No	Feature	Deskripsi
1.	Id	Nomor kolom
2.	Gender	Jenis kelamin
3.	Age	Usia
4.	Hypertension	Tekanan darah
5.	Heart Disease	Kelainan jantung dari lahir
6.	Ever Married	Status
7.	Work Type	Status pekerjaan
8.	Residence Type	Status wilayah
9.	Avg Glucose	Kadar gula darah (mg/dL)
10.	Bmi	Golongan berat badan
11.	Smoking Status	Status merokok
12.	Stroke	Penyakit pembuluh darah otak



Gambar 1. Tahapan Penelitian

2.2 Preprocessing

Data mining membutuhkan *preprocessing* sebelum dapat diproses. *Preprocessing* data merupakan jenis proses yang menjalankan data mentah sebagai persiapan untuk operasi pemrosesan lainnya. Tahap *preprocessing* dideskripsikan sebagai fase awal untuk membersihkan data yang tidak diinginkan atau tidak berarti [8]. Tujuan dari *preprocessing* pada data mining ialah bertujuan mengubah data ke dalam format yang membuat proses menjadi sederhana dan ampuh sesuai dengan harapan pengguna [9]. Serta indikasi untuk mendapatkan hasil yang lebih akurat dan mengurangi nilai data tanpa mengubah informasi yang terkandung. Terdapat berbagai proses yang dipakai untuk preprocessing seperti:

- a. *Cleaning* data
Proses menghilangkan noise dan data yang tidak relevan atau tidak konsisten disebut pembersihan data. Dalam hal ini, transaksi kurang dari dua item (satu item) dihilangkan [10].
- b. *Splitting* data
Membagi data menjadi dua bagian. Satu digunakan sebagai data pelatihan dan yang lainnya digunakan sebagai data uji. Dengan menggunakan data split, percobaan pelatihan dilakukan berdasarkan rasio split 70:30, dan rasio split dari data pelatihan yang tersisa dihitung sebagai data uji. Data latih ialah data akan digunakan untuk melakukan pelatihan, data uji ialah data yang belum pernah digunakan dalam pelatihan, dan berfungsi sebagai data untuk menguji keberhasilan atau keakuratan hasil pelatihan [11].

2.3 Algoritma Support Vector Regression (SVR)

Support Vector Regression (SVR) adalah kemajuan dalam *Support Vector Machine (SVM)* dalam hal regresi. *Support Vector Machine* merupakan metode klasifikasi yang mengkaji daerah-daerah yang memisahkan kategori dalam pengamatan [12]. SVR bertujuan untuk mencari fungsi $f(x)$ sebagai *hyperplane* (garis pembagi) berupa fungsi regresi yang cocok untuk semua data input dengan kesalahan sekecil mungkin [13].

Tujuan dari SVR ini adalah untuk memetakan vektor input ke dimensi yang lebih tinggi [14]. Persamaan umum *hyperplane* umum dapat dinyatakan sebagai berikut:

$$f(x) = w^T \varphi(x) + b \quad (1)$$

Dimana $\varphi(x)$ menunjukkan suatu titik didalam ruang fitur berdimensi lebih tinggi, hasil pemetaan dari input vektor x di dalam ruang input yang berdimensi lebih rendah. Koefisien w dan b diestimasi dengan cara meminimalkan fungsi resiko (*risk function*) yang didefinisikan dalam persamaan [18]:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\lambda} \sum_{i=1}^{\lambda} L_{\epsilon}(y_i, f(x_i)) \quad (2)$$

yang memenuhi:

$$\begin{aligned} y_i - w\varphi(x_i) - b &\leq \epsilon \\ w\varphi(x_i) - y_i + b &\leq \epsilon, i = 1, 2, \dots, \lambda \end{aligned} \quad (3)$$

Dimana,

$$L_{\epsilon}(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \epsilon & |y_i - f(x_i)| \geq \epsilon \\ 0 & \text{untuk yang lain} \end{cases} \quad (4)$$

Merupakan bobot dan b adalah perpotongan pada $x = 0$. Rentang toleransi direpresentasikan sebagai epsilon ϵ .

2.4 Algoritma Decision Tree C4.5

Algoritma pohon keputusan C4.5 adalah algoritma yang mengambil dataset berlabel dan menghasilkan pohon keputusan sebagai output. Pohon keputusan tindak lanjut ini kemudian divalidasi terhadap data uji yang ditandai dengan tidak terlihat untuk menghitung generalisasinya. Algoritma ini menggunakan konsep entropi informasi untuk membangun pohon keputusan dari sekumpulan data pelatihan, mirip dengan algoritma ID3 [15]. Pohon keputusan adalah metode yang sangat kuat dan terkenal untuk klasifikasi prediktif (pengelompokan). Metode pohon keputusan yang mewakili aturan mengubah menjadi fakta yang sangat besar. Aturannya mudah dimengerti dalam bahasa alami [16].

Adapun tahapan algoritma *Decision Tree C4.5* adalah sebagai berikut:

- Siapkan dataset pelatihan.
- Tentukan akar dari pohon keputusan.
- Hitung *gain*. Guna memilih atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai *Gain* dapat menggunakan persamaan satu sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^N \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (5)$$

- Prosedur untuk setiap cabang terpenuhi Ulangi langkah kedua. Di sisi lain, untuk menghitung nilai entropi, dapat menggunakan Persamaan dua di bawah ini:

$$Entropy(S) = \sum_{i=1}^N -\pi \times \log 2 \pi \quad (6)$$

- Proses partisipasi pohon keputusan berakhir ketika semua cabang node N mendapatkan kelas yang sama.

2.5 Evaluasi

Pada tahapan ini akan dilakukan evaluasi metode prediksi dengan mengukur performa terhadap algoritma *Support Vector Regression* dan *Decision Tree C4.5*. Evaluasi metode yang dilakukan ialah membandingkan nilai *error* masing-masing algoritma dengan menggunakan data yang sama dan *splitting* data yang sama yaitu 70:30.

3. HASIL DAN PEMBAHASAN

Setelah dilakukan pengujian dua algoritma, yaitu algoritma *Support Vector Regression* dan *Decision Tree C4.5* menghasilkan prediksi kasus pada pasien stroke berbeda. Pengujian ini menggunakan 5110 dataset dengan 12 *feature* dari dataset tersedia, Setiap bagian data diuji kepada semua data tersedia. Lalu mengukur hasil tes dan menghitung kesalahan untuk menentukan tingkat akurasi dalam memprediksi pasien stroke. Adapun dataset yang digunakan untuk menyusun laporan ini dapat dilihat pada tabel 2.

Tabel 2. Dataset yang digunakan

Id	Gender	Age	Hypertension	Heart Disease	Ever Married	Work Type	Residence Type	Avg Glucose	BMI	Smoking Status	Stroke
1	Male	67.0	0.0	1.0	Yes	Private	Urban	228.69	36.6	formerly smoked	Yes
2	Female	61.0	0.0	0.0	Yes	Self-employed	Rural	202.21	?	never smoked	Yes
3	Male	80.0	0.0	1.0	Yes	Private	Rural	105.92	32.5	never smoked	Yes
4	Female	49.0	0.0	0.0	Yes	Private	Urban	171.23	34.4	smokes	Yes
5	Female	79.0	1.0	0.0	Yes	Self-employed	Rural	174.12	24.0	never smoked	Yes
...
5110	Female	44.0	0.0	0.0	Yes	Govt_job	Urban	85.28	26.2	unknown	No

Sebelum data dianalisis, Pada proses ini dilakukan tahap preprocessing data yaitu penghapusan data yang bernilai kosong atau null dan transformasi data agar menyesuaikan dengan kebutuhan algoritma.. Adapun kode program yang memperlihatkan implementasi dari cleaning data dan transformasi data ditunjukkan pada tabel 3.

Tabel 3. Preprocessing Data

Id	Gender	Age	Hypertension	Heart Disease	Ever Married	Work Type	Residence Type	Avg Glucose	BMI	Smoking Status	Stroke
1	Male	67.0	0.0	1.0	Yes	Private	Urban	228.69	36.6	formerly smoked	No
2	Male	80.0	0.0	1.0	Yes	Private	Rural	105.92	32.5	never smoked	No
3	Female	79.0	1.0	0.0	Yes	Private	Rural	174.12	24.0	never smoked	No
4	Male	81.0	0.0	0.0	Yes	Private	Urban	186.21	29.0	formerly smoked	No
5	Male	74.0	1.0	1.0	Yes	Private	Rural	70.09	27.4	never smoked	No
6	Female	69.0	0.0	0.0	No	Govt_job	Urban	94.39	22.8	never smoked	No
7	Female	78.0	0.0	0.0	Yes	Private	Urban	58.57	24.2	unknown	Yes
...
3436	Female	44	0	0	Yes	Govt_job	Urban	85,28	26,2	Unknown	No

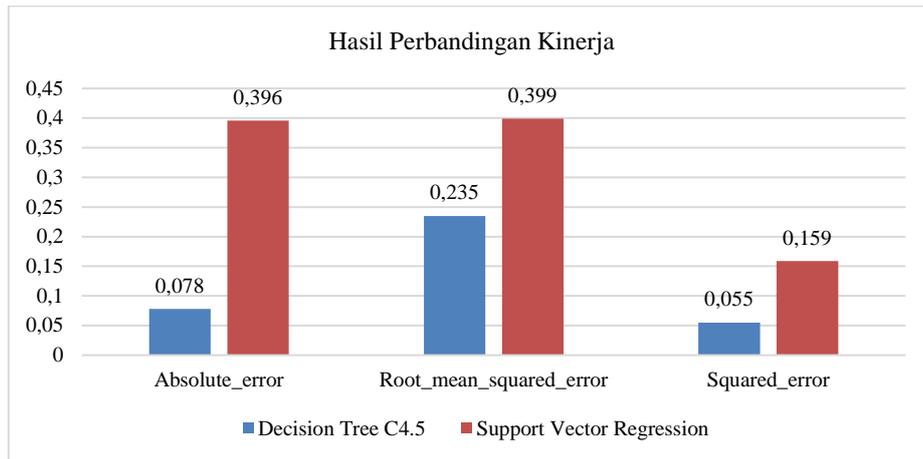
Setelah data melakukan tahap preprocessing ke dalam bentuk yang tepat, dataset tersebut berjumlah 5110 transaksi dengan 12 *feature*. Selanjutnya adalah melakukan Implementasi model algoritma *Support Vector Regression* dan *Decision Tree C4.5* yang telah dibangun dengan rasio 70:30. Berikut adalah Implementasi model algoritma *Support Vector Regression* dan *Decision Tree C4.5* yang telah dibangun dengan rasio 70:30 seperti pada tabel 4.

Tabel 4. Perbandingan Kinerja

Metode	Absolute_error	Root_mean_squared_error	Squared_error
<i>Decision Tree C4.5</i>	0.078 +/- 0.221	0.235 +/- 0.000	0.055 +/- 0.209
<i>Support Vector Regression</i>	0.396 +/- 0.045	0.399 +/- 0.000	0.159 +/- 0.044

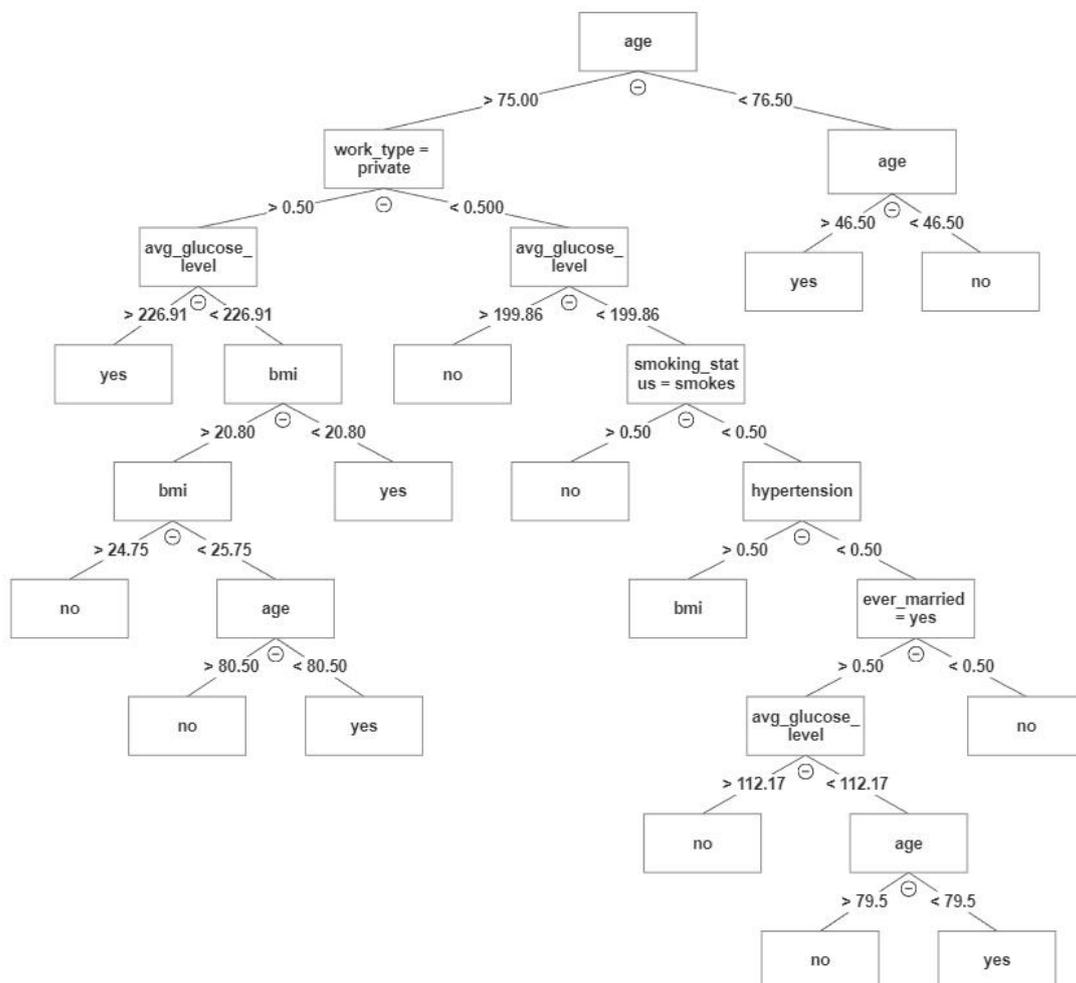
Gambar 2 menunjukkan hasil perbandingan algoritma *Support Vector Regression* dan *Decision Tree C4.5* yang telah dibangun dengan rasio 70:30 berdasarkan tingkatan *error*.

Pengujian algoritma di atas menunjukkan bahwa error yang dihasilkan oleh algoritma *Decision Tree C4.5* terhadap rasio 70 : 30 bernilai 0.235 dan algoritma *Support Vector Regression* terhadap rasio 70 : 30 bernilai 0.399. Saat menggunakan algoritma *Decision Tree C4.5*, maka akan menghasilkan output tambahan berupa diagram pohon keputusan dimana terdapat alur dalam memprediksi. Berikut adalah grafik pohon keputusan yang dihasilkan dengan rasio 70 : 30 dapat dilihat pada gambar 3.



Gambar 2. Hasil Diagram Perbandingan Kinerja

Pohon Keputusan Algoritma Decision Tree C4.5



Gambar 3. Pohon Keputusan Algoritma Decision Tree C4.5

4. KESIMPULAN

Hasil penelitian yang dilakukan dengan *splitting* data 70:30 terhadap 5110 dataset pasien penderita stroke yang diperoleh dari *kaggle open datasets* dengan menggunakan algoritma *Support Vector Regression* dan *Decision Tree C4.5* menyatakan bahwa algoritma *Decision Tree C4.5* sangat bagus dalam memprediksi dataset pasien penyakit stroke karena nilai *root mean square error* yang dihasilkan sebesar 0.235 dimana lebih

kecil dibandingkan hasil dari algoritma *Support Vector Regression* yang menghasilkan *root mean square error* sebesar 0.3999. Suatu algoritma dikatakan baik dan akurat dalam memprediksi data jika nilai *root mean square error* yang dihasilkan mendekati nilai 0 [17].

REFERENSI

- [1] S. Joko, *Data Mining : Algoritma dan Implementasi dengan Pemrograman PHP*, 2019.
- [2] F. Ramdhani dan A. Mutamakin, "Analisa Klasifikasi Biaya Pasien Rawat Inap Menggunakan Teknik Data Mining Attribute Importance (AI) dan Algoritma Naive Bayes".
- [3] KEMENKES RI, "Profil Kesehatan Indonesia 2020," 2021.
- [4] D. Haryadi, D. Marini Umi Atmaja, A. Rahman Hakim dan N. Suwaryo, "IDENTIFIKASI TINGKAT RESIKO PENYAKIT STROKE MENGGUNAKAN ALGORITMA REGRESI LINEAR BERGANDA," *Deny Haryadi, SNTEM*, vol. 1, pp. 1198-1207, 2021.
- [5] A. E. Pramadhani dan T. Setiadi, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PREDIKSI PENYAKIT ISPA (Infeksi Saluran Pernapasan Akut) DENGAN ALGORITMA DECISION TREE (ID3)," *Jurnal Sarjana Teknik Informatika*, vol. 2, no. 1, 2014.
- [6] Z. Rustam dan P. Kintandani, "Application of Support Vector Regression in Indonesian Stock," *Modelling and Simulation in Engineering*, vol. 2019, 2019.
- [7] S. Phatak, I. Mishra dan A. Swetapadma, "An Assessment of Decision Tree based Classification and Regression Algorithms," *2018 3rd International Conference on Inventive Computation Technologies (ICICT)*, pp. 92-95, 2018.
- [8] R. Rahmadden, M. K. Anam, Y. Irawan, S. Susandri dan M. Jamaris, "Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, 2022.
- [9] P. Meilina, "PENERAPAN DATA MINING DENGAN METODE KALSIFIKASI MENGGUNAKAN DECISION TREE DAN REGRESI," *Januari*, vol. 7, no. 1, 2015.
- [10] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *Jurnal Teknologi Rekayasa*, vol. 3, no. 1, p. 89, 2018.
- [11] I. Padiku, "Penerapan Metode Naive Bayes Classifier (Nbc) Untuk Klasifikasi Kondisi Internal Program Studi," *Jurnal Teknik*, vol. 19, no. 1, pp. 65-74, 2021.
- [12] E. S. R. Br.Situmorang, M. K. Anam, R. Rahmadden dan A. N. Ulfah, "Perbandingan Algoritma Svm Dan Nbc Dalam Analisa Sentimen Pilkada Pada Twitter," *CSRID (Computer Science Research and Its Development Journal)*, vol. 13, no. 3, 2021.
- [13] B. Scholkopf dan A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, 2006.
- [14] R. Amanda, H. Yasin dan A. Prahutama, "ANALISIS SUPPORT VECTOR REGRESSION (SVR) DALAM MEMPREDIKSI KURS RUPIAH TERHADAP DOLLAR AMERIKA SERIKAT," vol. 3, no. 4, pp. 849-857, 2014.
- [15] P. B. N. Setio, D. R. S. Saputro dan B. Winarno, "PRISMA, Prosiding Seminar Nasional Matematika Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," vol. 3, pp. 64-71, 2020.
- [16] A. Ginanjar Mabur dan R. Lubis, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI KRITERIA NASABAH KREDIT," *Jurnal Komputer dan Informatika (KOMPUTA) 53 Edisi. I*, no. 1, 2012.
- [17] I. Suprayogi, Trimaijon dan Mahyudin, "Model Prediksi Liku Kalibrasi Menggunakan Pendekatan Jaringan Saraf Tiruan (ZST)(Studi Kasus: Sub DAS Siak Hulu)," *Jurnal Online Mahasiswa Fakultas Teknik Universitas Riau*, vol. 1, no. 1, pp. 1-18, 2014.
- [18] M Mustakim, B Agus dan H Irman, "Performance Comparison Between Support Vector Regression and Artificial Neural Network for Prediction of Oil Palm Production", *Jurnal Ilmu Komputer dan Informasi*, vol. 9, no. 1, pp. 1-8, 2016.