

Implementasi Algoritma C4.5 dalam Melakukan Klasifikasi Penyakit Stroke Otak

Felian Nabila¹, Iis Afrianty^{2*}, Suwanto Sanjaya³, Fadhilah Syafria⁴

^{1,2,3,4}Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Panam, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau, Indonesia, 28293
e-mail: ¹11850124646@students.uin-suska.ac.id, ^{2*}iis.afrianty@uin-suska.ac.id,
³suwantosanjaya@uin-suska.ac.id, ⁴fadhilah.syafria@uin-suska.ac.id

Submitted Date: 2023-06-28
Revised Date: 2023-06-30

Reviewed Date: 2023-06-28
Accepted Date: 2023-06-30

Abstract

Stroke is a chronic health disorder and ranks among the diseases with the highest fatality rates. Rapid treatment of stroke results in a decreased incidence of complications and brain damage. Therefore, it is important to self-evaluate whether a person has suffered a stroke or not. This study we classified brain stroke disease using the C4.5 algorithm and a decision tree model to analyze data related to stroke disease, and divided the dataset into three sets: train set, validation set, and test set. The ratio is 70:20:10, then obtained result with a high accuracy of 95% in each data train set, validation set, test set. And then precision 0.91, recall 0.54, and f1-score of 0.56 for training set. for validation set, precision 0.48, than recall 0.50, and f1-score 0.49, while for the test set, the precision 0.48, recall 0.50, and f1-score 0.49. In summary, it can be inferred that the C4.5 decision tree algorithm demonstrates excellent performance in classifying stroke disease.

Keywords: Brain Stroke; C4.5; Data Mining; Decision Tree; Classification;

Abstrak

Stroke merupakan gangguan kesehatan dalam jangka panjang dan menjadi salah satu penyakit dengan resiko kematian paling tinggi. Penanganan stroke dengan cepat menyebabkan tingkat kemunculan komplikasi dan kerusakan yang terjadi pada otak berkurang. Oleh karena itu perlunya melakukan analisa diri pada orang yang bersangkutan apakah orang tersebut mengalami penyakit stroke atau tidak. Penelitian ini melakukan klasifikasi algoritma C4.5 penyakit *brain stroke* guna menganalisa data terkait penyebab stroke dengan model *decision tree* dan membagi dataset menjadi 3 yakni *train set*, *validation set*, *test set* dengan perbandingan 70:20:10, kemudian didapatkanlah hasil dengan akurasi yang tinggi sebesar 95% di setiap data *train set*, *validation set*, *test set*. Serta presisi sebesar 0,91, *recall* sebesar 0,54, *f1-score* sebesar 0,56 untuk data *train set*, kemudian presisi sebesar 0,48, *recall* sebesar 0,50, *f1-score* sebesar 0,49 untuk *validation set*, dan presisi sebesar 0,48, *recall* sebesar 0,50, *f1-score* sebesar 0,49 untuk *test set*. Dapat Disimpulkan bahwa algoritma C4.5 *decision tree* ini dapat melakukan klasifikasi penyakit stroke dengan sangat baik.

Kata Kunci: Brain Stroke; C4.5; Data Mining; Decision Tree; Klasifikasi

1. Pendahuluan

Stroke merupakan gangguan kesehatan dalam jangka panjang dan menjadi salah satu penyakit dengan resiko kematian paling tinggi. Terlebih lagi jika orang tersebut memiliki gangguan pada jantung, maka akan memperburuk kondisinya (S & T, 2020). Stroke terjadi saat

pasokan darah ke otak terhalang atau terganggu. Batang otak yaitu bagian dari struktur otak yang mengendalikan organ, serta anggota tubuh seperti detak jantung, tekanan darah, pernapasan. Stroke batang otak disebabkan oleh gangguan yang terjadi pada aliran darah di sekitar batang otak.

Penanganan stroke dengan cepat menyebabkan tingkat kemunculan komplikasi dan kerusakan yang terjadi pada otak berkurang. Deteksi penyakit stroke dalam beberapa jam pertama dapat meningkatkan peluang pencegahan komplikasi serta perawatan kesehatan. Efek yang signifikan dari jenis obat-obatan yang digunakan untuk pengobatan stroke hanya akan muncul pada tiga jam pertama sejak diberikan di awal stroke (Almadani & Alshammari, 2018). Oleh karena itu perlunya melakukan prediksi pada orang yang bersangkutan apakah orang tersebut mengalami penyakit stroke atau tidak. Cara untuk melakukan prediksi penyakit stroke pada seseorang salah satunya yaitu dengan klasifikasi guna memperoleh hasil yang akurat.

Klasifikasi adalah metode penambangan data yang dipergunakan untuk mengategorikan data ke dalam kelas atau grup yang telah ditentukan sebelumnya. Tujuan klasifikasi ialah untuk menemukan model yang dapat memberikan wawasan serta membedakan antara kelas ataupun konsep data yang berbeda. Tujuannya agar dapat dipergunakan untuk prediksi kelas dari objek yang tidak memiliki label kelas yang diketahui sebelumnya (Rafiska et al., 2018).

Pohon keputusan atau *decision tree* ialah metode yang bisa dipergunakan untuk melakukan pengelompokan atau klasifikasi penyakit stroke otak. *Decision tree* atau pohon keputusan melakukan klasifikasi sampel secara pendekatan *top-down* yang bermula dari simpul akar serta berlanjut secara sistematis sesuai dengan hasil tes pada simpul internal., hingga mencapai simpul daun, di mana label kelas diberikan kepada sampel tersebut (Pertiwi et al., 2019).

Algoritma C4.5 ialah algoritma yang dipergunakan dalam pembuatan *decision tree* atau pohon keputusan, yang mana algoritma ini banyak dipakai oleh para peneliti di dunia. Berdasarkan penjelasan tersebut, pada penelitian ini akan dilakukannya proses klasifikasi untuk penyakit stroke otak menerapkan algoritma C4.5 yang mana algoritma ini populer dan banyak digunakan secara praktis (Syamsul et al., 2018).

Penelitian sebelumnya terkait metode *data mining* untuk melakukan klasifikasi telah dilakukan oleh Pratama, Ridwan dan Prihandono (Pratama et al., 2021) pada penelitian tersebut digunakannya algoritma C4.5 guna melakukan klasifikasi menggunakan dua kelas yakni individu yang mengalami kanker serviks dan individu yang berada dalam kondisi sehat. Hasil yang diperoleh

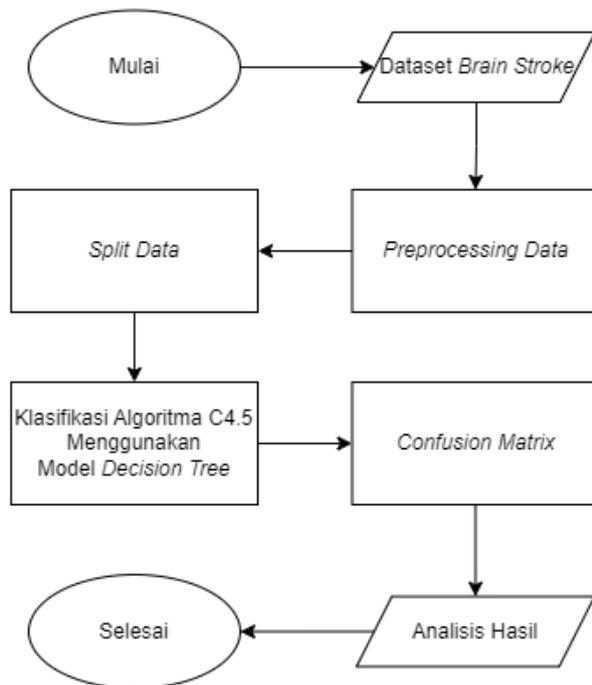
yaitu tingkat akurasi sebesar 98,61%, *precision* 98,08%, *recall/sensitivity* 95,24%, dan kurva *Receiver Operating Characteristic* (ROC) menampilkan angka 0,982% yang maknanya *excellent classification* atau baik.

Dalam penelitian lain yang dilakukan oleh Yunita dan Ikasari (Yunita & Ikasari, 2021) pada penelitian tersebut membandingkan kinerja dari dua algoritma yakni C4.5 dan Naïve Bayes guna melakukan tata cara klasifikasi tingkat kepuasan terhadap pelayanan yang diberikan oleh PT. Solusi Media Semesta. Hasil yang diperoleh yaitu tata cara dengan nilai akurasi tertinggi didapat oleh C4.5 sebesar 94,17%, sementara untuk Naïve Bayes akurasi yang didapat sebanyak 85,83%. Penelitian lainnya oleh Ardiansyah, Sunyoto dan Luthfi melakukan proporsi antara algoritma *naïve bayes* dengan algoritma C4.5 pada klasifikasi pada penyakit diabetes dengan 7 skenario berbeda. Didapatkanlah hasil yaitu algoritma C4.5 lebih tinggi pada tiap skenario daripada algoritma *naïve bayes* (Ardiansyah et al., 2021). Penelitian oleh Primajaya, Sari dan Khusaeri (Primajaya et al., 2020) melakukan prediksi potensi kebakaran hutan menggunakan C4.5 yang menghasilkan akurasi sebesar 89,7859%. Penelitian oleh Sari dan Yunus (Sari & Yunus, 2022) menggunakan metode C4.5 guna melakukan klasifikasi penyakit diabetes dan mendapatkan nilai akurasi 97,69%. Penelitian oleh Andarista dan Janato (Andarista & Jananto, 2022) melakukan klasifikasi dengan algoritma C4.5 untuk mengelompokkan kendaraan bermotor yang lulus uji dan tidak lulus uji menggunakan *tools Rapidminer*, dan menghasilkan akurasi 94,12%.

Dengan penelitian yang dilakukan diharapkan dapat memberikan pemahaman tentang penerapan data mining terkhususkah pada klasifikasi penyakit stroke otak. Penelitian ini berguna mengetahui nilai akurasi dan juga performa dari algoritma C4.5 dalam melakukan klasifikasi penyakit stroke otak.

2. Metodologi Penelitian

Dalam penelitian ini penggunaan metode algoritma C4.5 untuk klasifikasi dibagi menjadi beberapa tahapan proses agar dapat berjalan dengan semestinya. Berikut adalah gambaran mengenai tahapan penelitian yang dijelaskan sebagai berikut



Gambar 1 Flowchart Penelitian

Proses yang dilakukan pada penelitian ini dilihat pada Gambar 1 di atas. Pertama proses, digunakan dataset *brain stroke* yang bersumber dari *kaggle*, kemudian melakukan *preprocessing data*, lalu data dilakukan *split data* dengan perbandingan 70 : 20 : 10 untuk data *train*, *validation*, *test*. Setelah itu melakukan klasifikasi algoritma C4.5 dengan model *decision tree* dan menghitung *confusion matrix* serta menampilkan visualisasi *confusion matrix* dari tiap data *train*, *validation*, *test*. Lalu melakukan analisis hasil dengan data yang telah ada.

1) Dataset Brain Stroke

Data yang dipergunakan pada penelitian ini yaitu dataset *brain stroke* yang diperoleh dari platform *Kaggle* dengan jumlah sebanyak 4981 data. Dengan 10 variabel yakni *gender*, *age*, *hypertension*, *heart disease*, *ever married*, *work type*, *residence type*, *avg glucose level*, *bmi*, *smoking status*. Dan 1 *output* yakni *stroke*.

Tabel 1 Dataset Brain Stroke

Gender	Age	Hypertension	Heart disease	Ever married	Work type	Residence type	Avg Glucose Level	Bmi	Smoking Status	Stroke
--------	-----	--------------	---------------	--------------	-----------	----------------	-------------------	-----	----------------	--------

0	Male	67.0	0	1	Yes	Private	Urban	168.805	36.6	formerly smoked	yes
1	Male	80.0	0	1	Yes	Private	Rural	105.920	32.5	never smoked	yes
2	Female	49.0	0	0	Yes	Private	Urban	168.805	34.4	smokes	yes
3	Female	79.0	1	0	Yes	Self-employed	Rural	168.805	24.0	never smoked	yes
4	Male	81.0	0	0	Yes	Private	Urban	168.805	29.0	formerly smoked	yes
...
4976	Male	41.0	0	0	No	Private	Rural	70.150	29.8	formerly smoked	no
4977	Male	40.0	0	0	Yes	Private	Urban	168.805	31.1	smokes	no
4978	Female	45.0	1	0	Yes	Govt job	Rural	95.020	31.8	smokes	no
4979	Male	40.0	0	0	Yes	Private	Rural	83.940	30.0	smokes	no
4980	Female	80.0	1	0	Yes	Private	Urban	83.750	29.1	never smoked	no

Pada data tersebut, jumlah data yang terkena stroke sebanyak 248 data dan untuk yang tidak terkena stroke sebanyak 4733 data. Dari perbandingan jumlah data yang terkena stroke dan yang tidak terkena stroke, terdapat perbedaan yang sangat jauh.

2) Preprocessing Data

Dilakukannya *handling missing value* untuk mengelola nilai yang tidak lengkap atau hilang pada dataset. Kemudian melakukan *imbalance data* karena adanya ketimpangan pada data antara stroke dan tidak stroke, dilakukannya *imbalance data* menggunakan teknik *SMOTE* (*Synthetic Minority Over-sampling Technique*). Yang mana teknik ini membangkitkan data yang sedikit menjadi sebesar data yang paling banyak (Kasanah et al., 2019). Kemudian melakukan transformasi data yang akan mengubah variabel kategorikal menjadi numerikal agar dapat diproses, menggunakan metode *one-hot encoder*. Lalu dilakukannya normalisasi data karena adanya rentang nilai di tiap atribut dataset *brain stroke*

yang dapat membuat atribut dengan nilai kecil tidak berfungsi, diperlukan normalisasi data guna mengatasi hal tersebut (Nasution et al., 2019). Pada penelitian ini menggunakan normalisasi jenis min-max yang melakukan perubahan linier pada data asli, kemudian memberikan keseimbangan perbandingan tiap data. Normalisasi data menggunakan *Min - Max* dengan rumus:

$$V' = \frac{v - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A \quad (1)$$

Setelah selesai melakukan proses *handling missing value*, mengatasi *imbalance data*, serta transformasi data, lalu masuk proses selanjutnya. Proses tersebut yaitu melakukan implementasi normalisasi menggunakan *min-max*, dan mendapatkan hasil seperti pada Tabel 2 berikut

Tabel 2 Normalisasi

Age	Hypertension	Heart_disease	Avg_glucose_level	Bmi	Gender_Female	Gender_Male	Ever_married_No	Ever_married_Yes	Work_type	Residence_type	Smoking_status
0.523926	0.0	0.0	0.022368	0.338028	1.0	0.0	0.0	1.0	0.33333	1.0	1.00000
0.694824	0.0	0.0	1.00000	0.707355	1.0	0.0	0.0	1.0	0.33333	1.0	0.66667
0.890137	0.0	0.0	0.275813	0.453834	0.0	1.0	0.0	1.0	0.33333	0.0	0.33333
1.000000	0.0	0.0	0.013209	0.219092	0.0	1.0	0.0	1.0	0.66667	0.0	0.66667
0.389648	0.0	0.0	0.377262	0.350548	1.0	0.0	0.0	1.0	0.33333	1.0	1.00000

Penerapan normalisasi *min-max* untuk memastikan bahwa seluruh variabel memiliki pengaruh yang seimbang dalam prosesnya. Dan juga penerapan normalisasi ini dapat mengoptimalkan performa model klasifikasi yang dibangun.

3) Split Data

Setelah data selesai diolah pada tahapan *preprocessing data*, selanjutnya data tersebut dipisah menjadi 3 bagian yaitu *data train*, *data validation*, *data test*. Pada penelitian ini

menggunakan perbandingan 70 : 20 : 10. *Data train* yaitu bagian dataset yang digunakan untuk melatih model. Pada fase pelatihan, model akan disesuaikan dengan data ini guna mencari pola yang ada pada data sehingga dapat melakukan prediksi yang akurat. Untuk *data validation* digunakan untuk mengoptimalkan parameter-model selama proses pelatihan. Data ini tidak digunakan langsung untuk melatih model, tetapi digunakan untuk menguji performa model saat itu dan mengatur parameter agar performa model lebih baik. *Data validation* juga membantu memperbaiki masalah seperti *overfitting* atau *underfitting* pada model. Data test yaitu bagian dataset yang digunakan untuk menguji performa model yang telah dilatih dan disetel. Tujuan dari pengujian ini untuk mengevaluasi seberapa baik model yang dikembangkan bekerja pada data baru.

Pemisahan data menjadi tiga bagian ini penting untuk menghindari *overfitting* yang terjadi ketika model terlalu tertarik pada data pelatihan yang spesifik dan tidak dapat menggeneralisasi dengan baik pada data baru. Dengan membagi dataset menjadi *data train*, *data validation*, *data test*, kita dapat membangun dan mengoptimalkan model yang lebih akurat.

4) Klasifikasi Data Dengan Algoritma C4.5 model Decision Tree

Untuk melakukan klasifikasi dengan algoritma C4.5, digunakannya data latih yang telah melalui tahap *preprocessing data*. Kemudian tentukan atribut pada data latih guna sebagai kelas atau *node*, melakukan perhitungan *entropy* dan *gain* di tiap kelas, mencari nilai *gain* paling tinggi di tiap kelas dan membuat pohon keputusan atau *decision tree* berdasarkan kelas yang terpilih. Melakukan pengecekan di tiap kelas, jika kelas belum terisi maka lakukan perhitungan ulang pada nilai *entropy* dan *gain* sampai semua pohon terisi. Jika seluruh kelas sudah terisi, maka *decision tree* dapat dibangun.

Rumus menghitung *entropy* yaitu

$$Entropy(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (2)$$

S : Ruang Data (*Sample*)

n : Kemungkinan jumlah nilai pada atribut A (jumlah kelas)

p_i : Hasil pembagian antara jumlah tiap nilai dengan n



Rumus menghitung *gain* yaitu

$$\begin{aligned}
 & Gain(S, A) \\
 &= Entropy(S) \\
 &- \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (3)
 \end{aligned}$$

S : Himpunan Kasus

A : Atribut

|S_i| : Banyak kasus untuk nilai ke n

|S| : Banyak kasus yang ada didalam S

5) Hitung *Confusion Matrix*

Pada tahap ini dilakukannya evaluasi kinerja mode. Hasil dari *matrix* ini melakukan perbandingan dengan kelas aslinya (Badriah et al., 2021). Nilai akurasi di dalam *confusion matrix* mempunyai 2 kelas yang membentuk tabel *matrix* yaitu kelas positif dan negatif. *Confusion matrix* umumnya terdiri dari nilai:

- True Positive* (TP) yaitu banyak data yang secara benar diprediksi sebagai positif oleh model.
- True Negative* (TN) yaitu banyak data yang secara benar diprediksi sebagai negatif oleh model.
- False Positive* (FP) yaitu banyak data yang keliru diprediksi positif oleh model, yang sebenarnya negatif.
- False Negative* (FN) yaitu banyak data yang keliru diprediksi negatif oleh model, yang sebenarnya positif.

6) Analisis Hasil

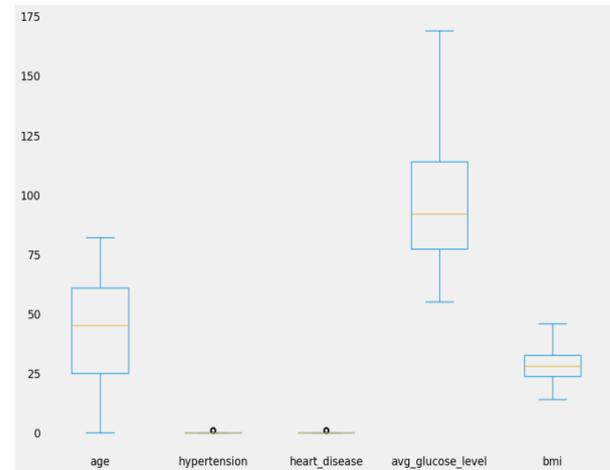
Pada tahap pengujian akan menghasilkan nilai akurasi, presisi, *recall* dan juga *f1score*, dari nilai tersebut dapat terlihat apakah performa algoritma C4.5 sangat baik terhadap pengklasifikasian penyakit stroke otak, melakukan perbandingan antara hasil penelitian dengan data yang ada serta penelitian lain dengan metode yang sama.

3. Hasil dan Pembahasan

Terdapat 4981 data *brain stroke* yang dipergunakan pada penelitian ini dengan 10 variabel. Hasil implementasi menggunakan *python google colaboratory*. Dilakukannya *handling missing value* untuk mengelola nilai yang tidak lengkap atau hilang pada dataset.

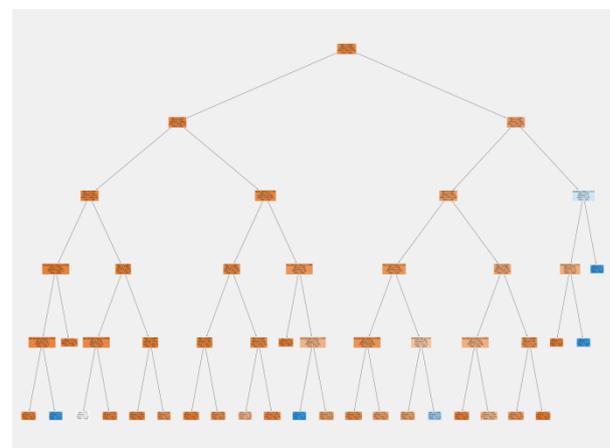
Melakukan *handling outlier* guna melakukan identifikasi nilai yang tidak biasa (*outlier*) pada dataset untuk mencegah hasil yang

tidak akurat saat analisis data. Berikut visualisasi *handling outlier*



Gambar 2 *Handling Outlier*

Kemudian, karena data yang timpang, maka dilakukannya *imbalance data* menggunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*). Setelah *imbalance data*, kemudian melakukan transformasi data menggunakan metode *one-hot encoder*, lalu dilakukan normalisasi menggunakan *minmax* (1) agar model bekerja lebih baik. Setelah itu, implementasi algoritma C4.5 (2)(3) pada data *brain stroke* dengan menerapkan model *decision Tree*. Dengan membagi 3 data yaitu data *train*, data *validation*, data *test* sebesar 70 : 20 : 10. Berikut merupakan hasil dengan model *decision tree* dari dataset yang telah diolah dengan menggunakan algoritma C4.5



Gambar 3 *Decision Tree* Algoritma C4.5

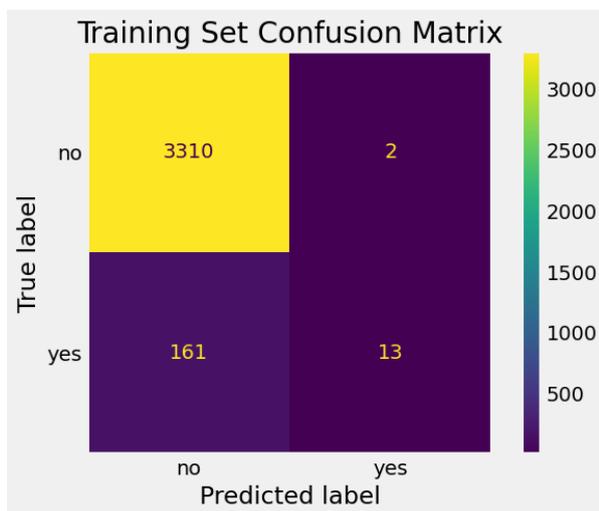
Penerapan algoritma C4.5 pada model *decision tree* ini mendapatkan hasil akurasi 95% di tiap data *train*, *validation*, *test*. Lalu presisi

sebesar 0,91, *recall* sebesar 0,54, *f1-score* sebesar 0,56 untuk data *train set*, kemudian presisi sebesar 0,48, *recall* sebesar 0,50, *f1-score* sebesar 0,49 untuk *validation set*, dan presisi sebesar 0,48, *recall* sebesar 0,50, *f1-score* sebesar 0,49 untuk *test set*.

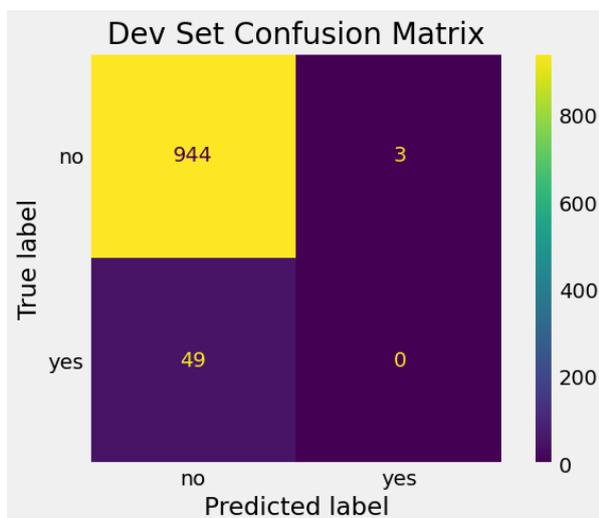
Tabel 3 Hasil Evaluasi dan *confusion matrix*

Data	Akurasi	Presisi	Recall	<i>f1-Score</i>
<i>Train Set</i>	95%	0,91	0,54	0,56
<i>Validation Set</i>	95%	0,48	0,50	0,49
<i>Test Set</i>	95%	0,48	0,50	0,49

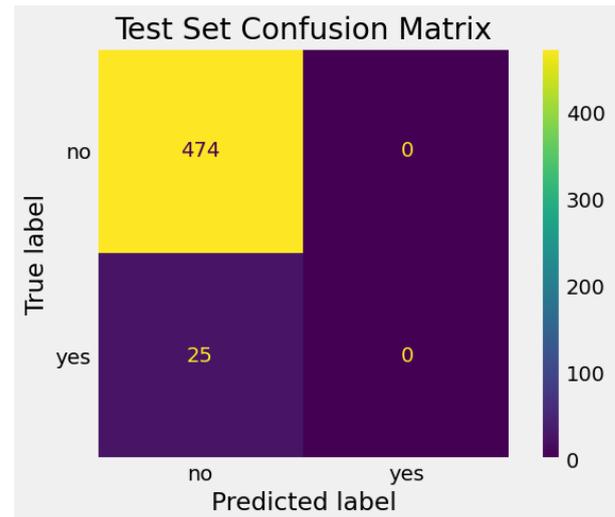
Berikut hasil *confusion matrix* seperti Gambar 4, Gambar 5, dan Gambar 6 berikut:



Gambar 4 *Confusion Matrix Training Set Data*



Gambar 5 *Confusion Matrix Validation Set Data*



Gambar 6 *Confusion Matrix Testing Set Data*

Confusion matrix membantu dalam mengukur performa model klasifikasi dengan memberikan informasi tentang ketepatan prediksi untuk setiap kelas. Dari *confusion matrix*, dapat menghitung berbagai matriks evaluasi seperti akurasi, presisi, *recall* dan *F1-score* untuk mengukur kinerja model secara lebih komprehensif.

4. Kesimpulan

Setelah dilakukannya penelitian terhadap data *brain stroke* dengan melakukan klasifikasi menggunakan algoritma C4.5 dengan model *decision tree*, kesimpulan yang dapat diambil dengan membagi dataset menjadi 3 yakni *train set*, *validation set*, *test set* dengan perbandingan 70:20:10, menghasilkan tingkat akurasi yang tinggi sebesar 95% di setiap data, serta presisi sebesar 0,91, *recall* sebesar 0,54, *f1-score* sebesar 0,56 untuk data *train set*, kemudian presisi sebesar 0,48, *recall* sebesar 0,50, *f1-score* sebesar 0,49 untuk *validation set*, dan presisi sebesar 0,48, *recall* sebesar 0,50, *f1-score* sebesar 0,49 untuk *test set*. Dapat Disimpulkan bahwa algoritma C4.5 *decision tree* ini dapat melakukan klasifikasi penyakit stroke dengan sangat baik.

5. Saran

Penelitian ini berhasil dalam melakukan implementasi algoritma C4.5 *decision tree* untuk klasifikasi penyakit *brain stroke*. Walau demikian, terdapat beberapa aspek yang dapat ditingkatkan untuk penelitian mendatang seperti melakukan uji coba terhadap beberapa perbandingan dengan rasio yang berbeda-beda untuk tiap data *train*, data *validation*, data *test*, membandingkan kinerja

algoritma C4.5 *decision tree* dengan beberapa algoritma lainnya untuk klasifikasi penyakit *brain stroke*, mengikutsertakan tenaga medis atau ahli terkait dalam menerjemahkan hasil klasifikasi penyakit stroke dapat memberikan sudut pandang medis yang lebih mendalam dan mengaitkan hasil klasifikasi dengan dampak medis yang lebih luas.

Daftar Pustaka

- Almadani, O., & Alshammari, R. (2018). Prediction of stroke using data mining classification techniques. *International Journal of Advanced Computer Science and Applications*, 9(1), 457–460.
<https://doi.org/10.14569/IJACSA.2018.090163>
- Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4. 5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, 16(2), 29–43.
<https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1525%0Ahttps://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/download/1525/944>
- Ardiansyah, M., Sunyoto, A., & Luthfi, E. T. (2021). Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 147–156.
<https://doi.org/10.29408/edumatic.v5i2.3424>
- Badriah, S., Nugroho, M. F. E., Sanjaya, N., Rismawati, I., Sari, B. N., & Rozikin, C. (2021). Klasifikasi Algoritma C4 . 5 dalam Menentukan Penerima Bantuan Covid-19 (Studi Kasus : Desa di Karawang). *JIP (Jurnal Informatika Polinema)*, 7(3), 23–28.
- Kasanah, A. N., Muladi, & Pujiyanto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 196–201.
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78.
<https://doi.org/10.24114/cess.v4i1.11458>
- Pertiwi, M. W., Adiwisastro, M. F., & Supriadi, D. (2019). Analisa Komparasi Menggunakan 5 Metode Data Mining dalam Klasifikasi Persentase Wanita Sudah menikah di Usia 15-49 yang Memakai Alat KB (Keluarga Berencana). *Jurnal Khatulistiwa Informatika*, 7(1), 37–42.
<https://doi.org/10.31294/jki.v7i1.5741>
- Pratama, T. G., Ridwan, A., & Prihandono, A. (2021). Penerapan Algoritma C4.5 untuk Klasifikasi Kanker Serviks Tingkat Awal. *Urecol Journal. Part E: Engineering*, 1(1), 1–6.
<https://doi.org/10.53017/uje.4>
- Primajaya, A., Sari, B. N., & Khusaeri, A. (2020). Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4.5 Studi Kasus Provinsi Kalimantan Barat. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 6(2), 188–192.
<https://doi.org/10.26418/jp.v6i2.37834>
- Rafiska, R., Defit, S., & Nurcahyo, G. W. (2018). Analisis Rekam Medis untuk Menentukan Pola Kelompok Penyakit Menggunakan Algoritma C4.5. 2(1), 391–396. <http://jurnal.iaii.or.id>
- S, M. B., & T, D. A. (2020). Predictive Model for Brain Stroke in CT using Deep Neural Network. *International Journal of Recent Technology and Engineering (IJRTE)*, 9(1), 2011–2017.
<https://doi.org/10.35940/ijrte.f9954.059120>
- Sari, A. N., & Yunus, R. (2022). Diabetes Classification Process Using C4.5 Method. *(APIT) Application of Information Technology*, 1(1), 7–11.
- Syamsul, B., Dwi, M., & Rahmi, H. (2018). Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, B24–B31.
- Yunita, D., & Ikasari, I. H. (2021). Perbandingan Metode Klasifikasi C4.5 dan Naïve Bayes untuk Mengukur Kepuasan Pelanggan. *Informatika Universitas Pamulang*, 6(3), 456–462.
<http://openjournal.unpam.ac.id/index.php/informatika456>