

Aplikasi prediksi resiko penyakit stroke menggunakan teknik *data mining*

Erfan Karyadiputra^{1*}, Agus Setiawan², Budi Ramadhani²

¹Program Studi Sistem Informasi

Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari
Jl. Adhyaksa No. 2, Banjarmasin, Kalimantan Selatan, Indonesia. 70104

*E-mail: erfantsy@gmail.com

²Program Studi Teknik Informatika

Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari
Jl. Adhyaksa No. 2, Banjarmasin, Kalimantan Selatan, Indonesia. 70104

Abstrak: Stroke merupakan salah satu penyakit yang seringkali memiliki dampak serius dan dapat mengancam nyawa. Deteksi dini risiko stroke memegang peran penting dalam pencegahan dan penanganan yang tepat. Salah satu teknik yang dapat digunakan untuk melakukan prediksi stroke yaitu menggunakan metode *data mining* yang dapat menganalisis pola data besar untuk membantu mengidentifikasi hubungan kompleks antara faktor risiko dan kemungkinan awal terjadinya stroke. Tujuan penelitian ini adalah untuk merancang aplikasi yang dapat memprediksi risiko penyakit stroke pada seseorang dengan memanfaatkan teknik *data mining*. Penelitian ini menggunakan metode *decision tree algorithm* untuk analisis data dan metode *waterfall* untuk perancangan sistem aplikasinya. Hasil penelitian menunjukkan bahwa aplikasi ini mampu melakukan prediksi penyakit stroke dengan akurasi algoritma sebesar 95,03%

Kata Kunci: aplikasi, *decision tree*, stroke, *waterfall*

Abstract: Stroke is a disease that often has serious and life-threatening effects. Early detection of stroke risk plays an important role in prevention and proper treatment. One of the techniques that can be used to predict stroke is using data mining methods that can analyse big data patterns to help identify complex relationships between risk factors and the initial likelihood of stroke. The purpose of this research is to design an application that can predict the risk of stroke in a person by utilising data mining techniques. This research uses the decision tree algorithm method for data analysis and the waterfall method for designing the application system. The results showed that this application was able to predict stroke disease with an algorithm accuracy of 95.03%.

Keywords: application, *decision tree*, stroke, *waterfall*

PENDAHULUAN

Stroke merupakan salah satu penyakit yang dapat menyebabkan dampak serius pada kesehatan dengan tingkat kematian tertinggi kedua di seluruh dunia (Werdingisih et al., 2023). Risiko terkena penyakit ini semakin meningkat, apabila seseorang juga memiliki masalah pada jantung, yang semakin memperburuk kondisinya (Purnamasari & Suprpti, 2024). Oleh karena itu, deteksi dini dan prediksi risiko stroke menjadi sangat penting untuk mengambil langkah-langkah pencegahan yang tepat (Permana et al., 2024). Seiring dengan perubahan gaya hidup, urbanisasi, dan peningkatan usia penduduk, terjadi peningkatan risiko faktor-faktor penyebab stroke seperti hipertensi, diabetes, merokok, dan pola makan tidak sehat (Azhar et al., 2022). Selain itu, akses terbatas terhadap layanan kesehatan dan edukasi kesehatan yang kurang optimal dapat memperburuk situasi (Pambudi et al., 2022). Faktor geografis dan sosial-ekonomi

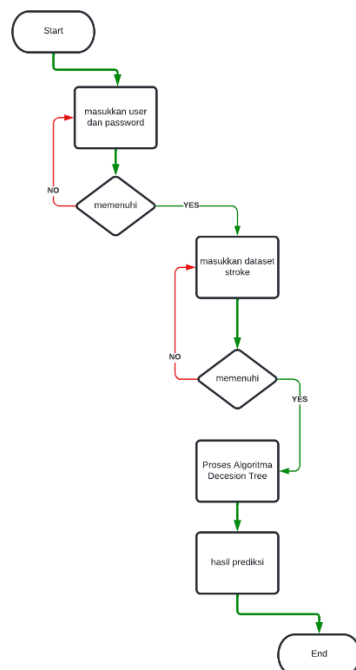
juga dapat berperan dalam kesenjangan akses layanan dan deteksi dini penyakit stroke (Sidiq et al., 2024).

Berdasarkan hasil analisis Dinas Kesehatan Provinsi Kalimantan Selatan menunjukkan bahwa prevalensi stroke meningkat seiring bertambahnya usia, dan rata-rata jenis kelamin didominasi perempuan, diikuti dengan pendidikan rendah dan tinggal di pedesaan serta orang yang tidak bekerja juga rentan terkena stroke. Dengan pertumbuhan jumlah kasus stroke yang semakin meningkat, perlu upaya penelitian dan intervensi kesehatan yang lebih intensif untuk merinci prevalensi, serta profil risiko individu yang beresiko terkena stroke (Puspitasari et al., 2021). Latar belakang ini menjadi dasar penting untuk merancang program pencegahan dan penanganan yang lebih tepat sasaran guna mengurangi beban penyakit stroke di kalangan masyarakat Kalimantan Selatan.

Dalam penelitian ini, digunakan metode data *mining* yaitu *decision tree*, yang mencakup analisis pola data besar untuk mengidentifikasi hubungan kompleks antara faktor risiko dan kemungkinan terjadinya stroke (Zuama et al., 2022). Selain itu, penelitian ini juga memanfaatkan algoritma data *mining* untuk membangun model prediksi yang akurat (Sofyan et al., 2023). Dengan menggabungkan informasi klinis dan data pasien, diharapkan aplikasi ini dapat memberikan kontribusi signifikan dalam membantu para profesional medis untuk melakukan tindakan pencegahan dan intervensi lebih awal pada individu dengan potensi risiko stroke. Penelitian ini diharapkan dapat memberikan kontribusi positif dalam pengembangan solusi kesehatan yang proaktif dan efektif.

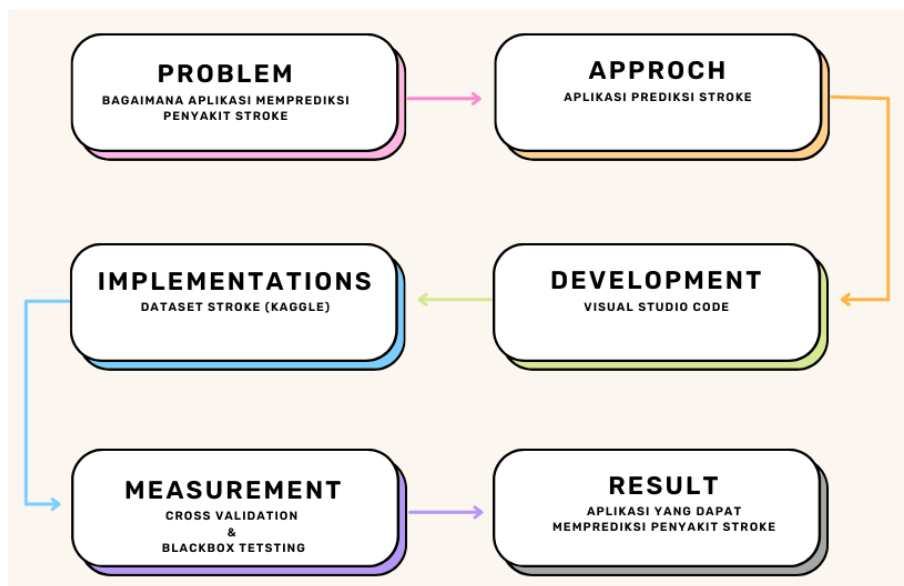
METODE PENELITIAN

Metode yang digunakan untuk membangun aplikasi prediksi stroke menggunakan metode *decision tree* untuk analisis data (Hermawan et al., 2022) dan metode *waterfall* untuk pengembangan aplikasinya dengan tahapan sebagaimana ditunjukkan pada Gambar 1 (Karyadiputra et al., 2023).



Gambar 1. Flowchart penelitian

1. Analisis kebutuhan
Pada tahap ini dilakukan analisis kebutuhan dengan cara mengidentifikasi kebutuhan fungsional dan non fungsional dari sistem yang akan dikembangkan.
2. Desain
Pada fase ini dilakukan perancangan sistem yang akan dibangun, meliputi arsitektur, struktur data, algoritma, dan antarmuka pengguna.
3. Implementasi
Tahap ini merupakan tahap dimana sistem dibuat secara fisik dengan menggunakan beberapa bahasa pemrograman.
4. Pengujian
Setelah implementasi selesai, sistem diuji untuk memastikan kinerjanya memenuhi spesifikasi yang ditetapkan.
5. Deployment
Setelah sistem lulus pengujian, langkah selanjutnya adalah implementasi seluruh sistem, termasuk instalasi dan konfigurasi perangkat lunak.
6. Pemeliharaan
Tahap selanjutnya melakukan pemeliharaan dan perbaikan yang diperlukan untuk memastikan sistem berfungsi dengan baik dan sesuai dengan kebutuhan pengguna.
Adapun prosedur penelitian berupa kerangka pemikiran dapat dilihat pada Gambar 2.



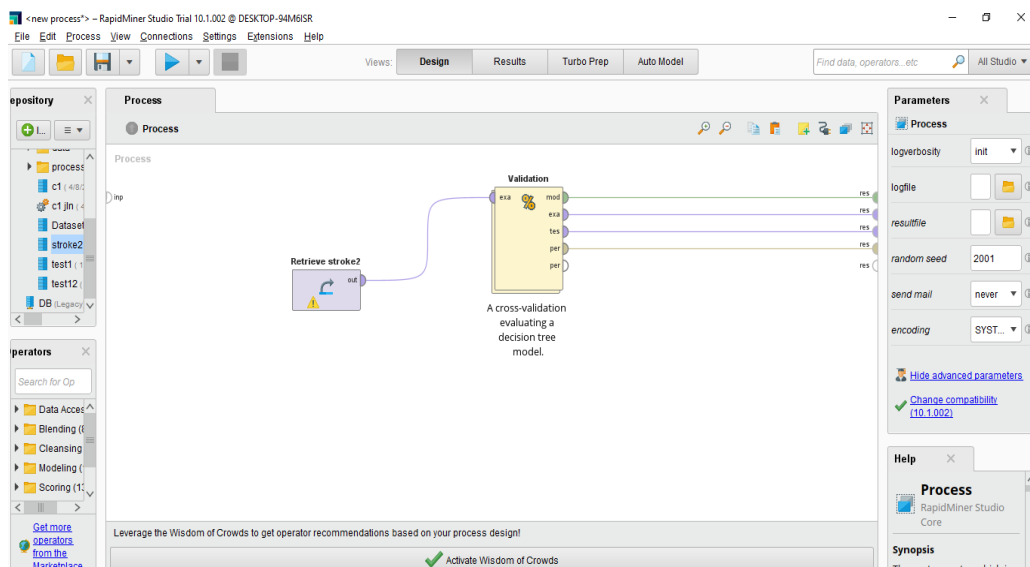
Gambar 2. Kerangka pemikiran

HASIL DAN PEMBAHASAN

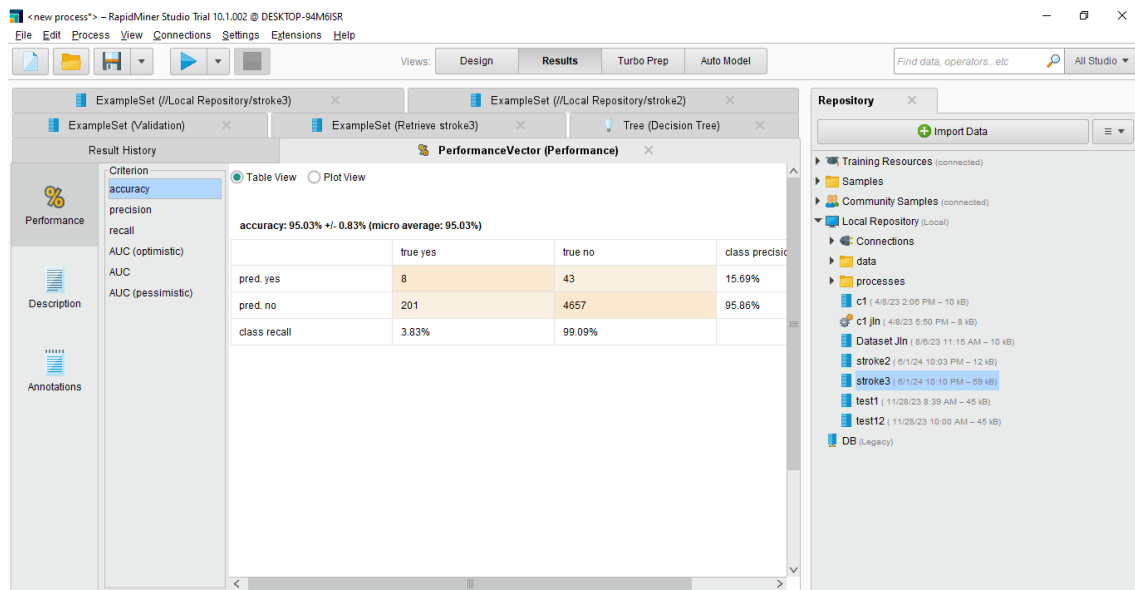
Pada penelitian ini menggunakan data stroke dari data publik Kaggle. Karakteristik data terdiri dari 11 variabel yaitu 10 sebagai variabel dependen (atribut x) dan 1 sebagai variabel independen (kelas y) dengan jumlah data sebanyak 4909 data. Setelah melakukan *preprocessing*, selanjutnya menguji algoritma *decision tree* menggunakan *cross validation*. *Cross validation* adalah teknik statistik yang digunakan dalam *machine learning* dan pemodelan prediktif lainnya untuk menilai kinerja dan kemampuan generalisasi suatu model (Witten & Frank, 2018).

Tabel 1. Fitur dataset stroke

No	Variabel	Nama	Keterangan
1	Y	Stroke	Yes No
2	X1	Gender (Jenis kelamin)	Male (Pria) Female (Wanita)
3	X2	Age (Umur)	Senior (Lansia 65 tahun ke atas) Older adult (Dewasa tua (56-65 tahun)) Middle adult (Dewasa tengah 36-55 tahun) Young adult (Dewasa muda 19-35 tahun) Teenager (Remaja 13-18 tahun)
4	X3	Hypertension (Tekanan darah tinggi)	Yes No
5	X4	Heart disease (Penyakit jantung)	Yes No
6	X5	Ever married (Pernah menikah)	Yes No
7	X6	Work type (Jenis pekerjaan)	Never worked (Tidak pernah bekerja) Govt job (Pekerja pemerintah) Private (Swasta) Self employed (Wiraswasta)
8	X7	Residence type (Jenis tempat tinggal)	Urban (Perkotaan) Rural (Pedesaan)
9	X8	Avg glucose level (Kadar glukosa)	Hipoglikemia <70 mg/dL Normal <100 mg/dL Prediabetes 100-125 mg/dL Diabetes ≥126 mg/dL
10	X9	Body Mass Index (BMI)	Underweight (Kurus <18,5) Normal weight (Normal <25) Overweight (Kelebihan berat badan <30) Obesity class 1 (Obesitas kelas 1 <35) Obesity class 2 (Obesitas kelas 2 <40) Obesity class 3 (Obesitas kelas 1 ≥ 40)
11	X10	Smoking status (Status merokok)	Smokes (Merokok) Formerly smoked (Sebelumnya merokok) Never smoked (Tidak pernah merokok) Unknown (Tidak diketahui)

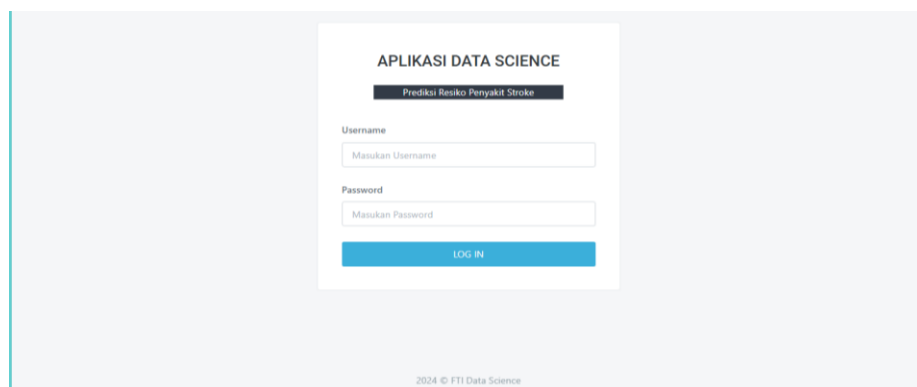


Gambar 3. Proses cross validation

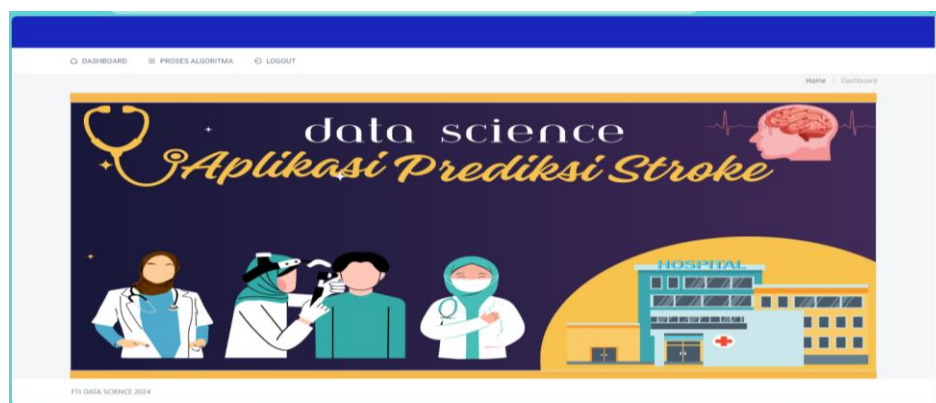


Gambar 4. Hasil akurasi

Hasil dari pengujian menggunakan *cross validation* didapatkan akurasi dalam memprediksi awal kemungkinan seseorang terindikasi stroke sebesar 95,03%. Berdasarkan hasil pengujian tersebut maka algoritma *decision tree* akan diimplementasikan ke dalam sebuah aplikasi sehingga memudahkan melakukan prediksi penyakit stroke (Putri et al., 2021).



Gambar 5. Halaman login



Gambar 6. Halaman utama aplikasi

Dashboard
PROSES ALGORITMA
Logout

Metode Perhitungan Decision Tree

Input Dataset
Tabel Dataset
Prediksi

Pilih Data Excel

Dengan format: .xlsx

No file chosen

Dataset Pertandingan

gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status
Male	Senior	no	yes	Yes	Private	Urban	Diabetes	Obesity_Class 2	formerly_smoked
Male	Senior	no	yes	Yes	Private	Rural	Prediabetes	Obesity_Class 1	never_smoked
Female	Middle_Adult	no	no	Yes	Private	Urban	Diabetes	Obesity_Class 1	smokes
Female	Senior	yes	no	Yes	Self-employed	Rural	Diabetes	Normal_weight	never_smoked
Male	Senior	no	no	Yes	Private	Urban	Diabetes	Overweight	formerly_smoked

FTI DATA SCIENCE 2024

Gambar 7. *Form upload dataset*

The screenshot shows a web application for Decision Tree calculations. The interface is clean and modern, with a light gray background. The left sidebar is a vertical panel with a white background, containing the application title and a list of input/output types. The main content area is a large white rectangle with a gray border, containing the 'Input Data Prediction' form. The form has 14 input fields, each with a dropdown menu. The fields are labeled with their respective variable names. The 'Predict' button is a green rectangle with white text, located at the bottom of the form.

Metode Perhitungan Decision Tree

Input Dataset
Target Dataset
Resultat

Input Data Prediction

gender
male

age
30

pekerjaan
manajemen

no
10

heart_disease
no

ever_married
married

sex
male

work_type
Private

residence_type
Urban

ang_glasses (none)

smoking
smoker

sex
female

Clarity_Class.2
Clarity_Class.2

smoking_status
formerly,smoked

Predict

Gambar 8. *Form input* data baru untuk prediksi

```

Decision Tree
1 0
2 0
3 0
4 0
5 0
6 0
7 0
8 0
9 0
10 0
11 0
12 0
13 0
14 0
15 0
16 0
17 0
18 0
19 0
20 0
21 0
22 0
23 0
24 0
25 0
26 0
27 0
28 0
29 0
30 0
31 0
32 0
33 0
34 0
35 0
36 0
37 0
38 0
39 0
40 0
41 0
42 0
43 0
44 0
45 0
46 0
47 0
48 0
49 0
50 0
51 0
52 0
53 0
54 0
55 0
56 0
57 0
58 0
59 0
60 0
61 0
62 0
63 0
64 0
65 0
66 0
67 0
68 0
69 0
70 0
71 0
72 0
73 0
74 0
75 0
76 0
77 0
78 0
79 0
80 0
81 0
82 0
83 0
84 0
85 0
86 0
87 0
88 0
89 0
90 0
91 0
92 0
93 0
94 0
95 0
96 0
97 0
98 0
99 0
100 0
101 0
102 0
103 0
104 0
105 0
106 0
107 0
108 0
109 0
110 0
111 0
112 0
113 0
114 0
115 0
116 0
117 0
118 0
119 0
120 0
121 0
122 0
123 0
124 0
125 0
126 0
127 0
128 0
129 0
130 0
131 0
132 0
133 0
134 0
135 0
136 0
137 0
138 0
139 0
140 0
141 0
142 0
143 0
144 0
145 0
146 0
147 0
148 0
149 0
150 0
151 0
152 0
153 0
154 0
155 0
156 0
157 0
158 0
159 0
160 0
161 0
162 0
163 0
164 0
165 0
166 0
167 0
168 0
169 0
170 0
171 0
172 0
173 0
174 0
175 0
176 0
177 0
178 0
179 0
180 0
181 0
182 0
183 0
184 0
185 0
186 0
187 0
188 0
189 0
190 0
191 0
192 0
193 0
194 0
195 0
196 0
197 0
198 0
199 0
200 0
201 0
202 0
203 0
204 0
205 0
206 0
207 0
208 0
209 0
210 0
211 0
212 0
213 0
214 0
215 0
216 0
217 0
218 0
219 0
220 0
221 0
222 0
223 0
224 0
225 0
226 0
227 0
228 0
229 0
230 0
231 0
232 0
233 0
234 0
235 0
236 0
237 0
238 0
239 0
240 0
241 0
242 0
243 0
244 0
245 0
246 0
247 0
248 0
249 0
250 0
251 0
252 0
253 0
254 0
255 0
256 0
257 0
258 0
259 0
260 0
261 0
262 0
263 0
264 0
265 0
266 0
267 0
268 0
269 0
270 0
271 0
272 0
273 0
274 0
275 0
276 0
277 0
278 0
279 0
280 0
281 0
282 0
283 0
284 0
285 0
286 0
287 0
288 0
289 0
290 0
291 0
292 0
293 0
294 0
295 0
296 0
297 0
298 0
299 0
300 0
301 0
302 0
303 0
304 0
305 0
306 0
307 0
308 0
309 0
310 0
311 0
312 0
313 0
314 0
315 0
316 0
317 0
318 0
319 0
320 0
321 0
322 0
323 0
324 0
325 0
326 0
327 0
328 0
329 0
330 0
331 0
332 0
333 0
334 0
335 0
336 0
337 0
338 0
339 0
340 0
341 0
342 0
343 0
344 0
345 0
346 0
347 0
348 0
349 0
350 0
351 0
352 0
353 0
354 0
355 0
356 0
357 0
358 0
359 0
360 0
361 0
362 0
363 0
364 0
365 0
366 0
367 0
368 0
369 0
370 0
371 0
372 0
373 0
374 0
375 0
376 0
377 0
378 0
379 0
380 0
381 0
382 0
383 0
384 0
385 0
386 0
387 0
388 0
389 0
390 0
391 0
392 0
393 0
394 0
395 0
396 0
397 0
398 0
399 0
400 0
401 0
402 0
403 0
404 0
405 0
406 0
407 0
408 0
409 0
410 0
411 0
412 0
413 0
414 0
415 0
416 0
417 0
418 0
419 0
420 0
421 0
422 0
423 0
424 0
425 0
426 0
427 0
428 0
429 0
430 0
431 0
432 0
433 0
434 0
435 0
436 0
437 0
438 0
439 0
440 0
441 0
442 0
443 0
444 0
445 0
446 0
447 0
448 0
449 0
450 0
451 0
452 0
453 0
454 0
455 0
456 0
457 0
458 0
459 0
460 0
461 0
462 0
463 0
464 0
465 0
466 0
467 0
468 0
469 0
470 0
471 0
472 0
473 0
474 0
475 0
476 0
477 0
478 0
479 0
480 0
481 0
482 0
483 0
484 0
485 0
486 0
487 0
488 0
489 0
490 0
491 0
492 0
493 0
494 0
495 0
496 0
497 0
498 0
499 0
500 0
501 0
502 0
503 0
504 0
505 0
506 0
507 0
508 0
509 0
510 0
511 0
512 0
513 0
514 0
515 0
516 0
517 0
518 0
519 0
520 0
521 0
522 0
523 0
524 0
525 0
526 0
527 0
528 0
529 0
530 0
531 0
532 0
533 0
534 0
535 0
536 0
537 0
538 0
539 0
540 0
541 0
542 0
543 0
544 0
545 0
546 0
547 0
548 0
549 0
550 0
551 0
552 0
553 0
554 0
555 0
556 0
557 0
558 0
559 0
560 0
561 0
562 0
563 0
564 0
565 0
566 0
567 0
568 0
569 0
570 0
571 0
572 0
573 0
574 0
575 0
576 0
577 0
578 0
579 0
580 0
581 0
582 0
583 0
584 0
585 0
586 0
587 0
588 0
589 0
590 0
591 0
592 0
593 0
594 0
595 0
596 0
597 0
598 0
599 0
600 0
601 0
602 0
603 0
604 0
605 0
606 0
607 0
608 0
609 0
610 0
611 0
612 0
613 0
614 0
615 0
616 0
617 0
618 0
619 0
620 0
621 0
622 0
623 0
624 0
625 0
626 0
627 0
628 0
629 0
630 0
631 0
632 0
633 0
634 0
635 0
636 0
637 0
638 0
639 0
640 0
641 0
642 0
643 0
644 0
645 0
646 0
647 0
648 0
649 0
650 0
651 0
652 0
653 0
654 0
655 0
656 0
657 0
658 0
659 0
660 0
661 0
662 0
663 0
664 0
665 0
666 0
667 0
668 0
669 0
670 0
671 0
672 0
673 0
674 0
675 0
676 0
677 0
678 0
679 0
680 0
681 0
682 0
683 0
684 0
685 0
686 0
687 0
688 0
689 0
690 0
691 0
692 0
693 0
694 0
695 0
696 0
697 0
698 0
699 0
700 
```

Gambar 9. *Form* hasil prediksi

KESIMPULAN

Stroke merupakan penyakit serius dengan dampak kesehatan yang signifikan dan tingkat kematian yang tinggi di seluruh dunia. Deteksi dini risiko stroke sangat penting untuk pencegahan dan penanganan yang tepat. Penelitian ini berfokus pada pengembangan aplikasi untuk memprediksi risiko stroke menggunakan teknik data *mining*, khususnya algoritma *decision tree*. Hasil penelitian menunjukkan bahwa aplikasi ini memiliki akurasi prediksi sebesar 95,03%. Penelitian juga menyoroti faktor-faktor risiko stroke seperti hipertensi, diabetes, merokok, dan pola makan tidak sehat, serta keterbatasan akses terhadap layanan kesehatan dan edukasi yang memperburuk situasi. Penelitian ini diharapkan dapat memberikan kontribusi dalam pencegahan dan penanganan stroke yang lebih efektif, terutama di kalangan masyarakat Kalimantan Selatan, dengan memperhatikan profil risiko individu yang rentan terhadap stroke.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada tim peneliti Data Science Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin. Penelitian ini didukung dan disponsori oleh Yayasan Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin.

DAFTAR PUSTAKA

- Anggraeni, D., & Ramadhani. (2018). Analisa *decision tree* untuk prediksi diagnosa diabetes mellitus. *Seminar Nasional Royal (SENAR)*, 1(1).
- Permana, B. A. C., & Patwari, I. K. D. (2021). Komparasi metode klasifikasi data *mining decision tree* dan *naïve bayes* untuk prediksi penyakit diabetes. *Infotek : Jurnal Informatika dan Teknologi*, 4(1), 63–69. <https://doi.org/10.29408/jit.v4i1.2994>.
- Detikhealth. (2019). *Mengenai Penyakit Diabetes, Penyebab dan Cara Mengatasinya*. <https://health.detik.com/berita-detikhealth/d-4468242/mengenai-penyakit-diabetes-penyebab-dan-cara-mengatasinya>.
- Edy. (2012). Penerapan algoritma c4.5 dengan seleksi atribut berbasis algoritma genetika dalam diagnosa penyakit jantung. *Tesis*. Jakarta: Sekolah Tinggi Manajemen Informatika dan Komputer Eresha.
- Eldridge, B. L. (2022). *The Significance of Diagnosing a Disease as Asymptomatic*. Verywell Health. <https://www.verywellhealth.com/asymptomatic-definition-importance-and-controversy-2249055>.
- Karyadiputra, E., & Hijriana, N. (2019). Penerapan algoritma *decision tree* C4.5 untuk klasifikasi penentuan daftar prioritas pengembangan jembatan. *Technologia*, 10(1), 43–46.
- Febrian, F. (2011). Algoritma klasifikasi data mining pada akseptasi data fakultatif. *Skripsi*. Jakarta: Sekolah Tinggi Manajemen Informatika dan Komputer Eresha Jakarta.
- Hastuti, K. (1994). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. *Semantik*, 2(1), 241–249. <https://doi.org/10.2307/j.ctv11hppt6.3>.
- Witten, I. H., & Frank, M. A. H. E. (2018). Data mining: Practical machine learning tools and techniques. *Angewandte Chemie International Edition*, 6(11), 951–952.
- International Diabetes Federation. (2021). *Diabetes Facts and Figures*. <https://www.idf.org/aboutdiabetes/what-is-diabetes/glossary.html>.
- Islam, M. M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. *Advances in Intelligent Systems and Computing*, 992, 113–125. https://doi.org/10.1007/978-981-13-8798-2_12.
- Kementerian Kesehatan RI. (2021). *Penyakit Diabetes Melitus*. <http://p2ptm.kemkes.go.id/infographic-p2ptm/penyakit-diabetes-melitus>.
- Najib, A., Nurcahyono, D., & Setiawan, R. P. P. (2019). Klasifikasi diagnosa penyakit diabetes mellitus (DM) menggunakan algoritma C4.5. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 11(2), 47–58. <https://doi.org/10.46964/justti.v11i2.153>.
- Putri, S. U., Irawan, E., & Rizky, F. (2021). Implementasi data mining untuk prediksi penyakit diabetes dengan algoritma C4.5. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 2(1), 39–46. <https://doi.org/10.30645/kesatria.v2i1.56>.

- Rahman, F. Y. (2021). Penerapan metode waterfall pada aplikasi laundry berbasis web. *Technologia: Jurnal Ilmiah*, 12(2), 125–132.
- Rifqi, M. (2016). Aplikasi data mining untuk diagnosis penyakit diabetes menggunakan algoritma C4.5 dan naïve bayes classification. *Skripsi*. Semarang: Universitas Dipenogoro.
- Rumini, & Nasruddin, A. (2021). Prediksi awal penyakit diabetes mellitus menggunakan algoritma naïve bayes. *Jurnal ICT: Information. Communication & Technology*, 20(2), 246–253.
- Suradiradja, K. H. (2012). Deteksi transaksi pencucian uang dengan algoritma klasifikasi C4.5. *Jurnal ESIT(E-Bisnis, Sistem Informasi, Teknologi Informasi)*, 12(1), 62-66.
- .