

Klasifikasi Risiko Strok Menggunakan Algoritma Random Forest dengan Teknik Knowledge Discovery in Database

Deni Athallah Ubaid¹, Fathoni², Muhammad Ichsan Farel Rachmad³

Universitas Sriwijaya¹²³

09031382227148@student.unsri.ac.id¹, fathoni@unsri.ac.id², 09031382227151@student.unsri.ac.id³

Diterima (31-03-2025)	Direvisi (19-04-2025)	Disetujui (24-04-2025)
--------------------------	--------------------------	---------------------------

Abstrak - Strok merupakan suatu kondisi ketika aliran darah ke otak mengalami gangguan, yang dapat menyebabkan kerusakan otak bahkan kematian. Strok juga merupakan penyebab utama kematian dan kecacatan yang terjadi pada seluruh dunia. Dengan tingkat kecelakaan yang tinggi, deteksi dini strok sangat diperlukan untuk mengurangi risiko strok. Penelitian ini bertujuan untuk mengklasifikasi risiko strok berdasarkan algoritma *Random Forest* dengan melalui pendekatan *Knowledge Discovery in Database* (KDD). Dataset yang digunakan merupakan dataset dari platform Kaggle yang terdiri dari 35.000 data pasien dengan 18 atribut relevan yang telah melalui proses pembersihan dan fitur pemilihan untuk mencapai akurasi prediksi yang optimal. Model training dilakukan dengan membagi data sebesar 80% data training dan 20% data testing. Hasil dari evaluasi menggunakan confusion matrixnya menunjukkan bahwa Algoritma Random Forest memiliki akurasi sebesar 87,47% dengan precision pada kelas 1 sebesar 96,40% dan recall sebesar 68,53%, serta precision pada kelas 0 sebesar 84,51% dan recall sebesar 98,51%. Namun, tantangan utama dalam penelitian ini adalah ketidakseimbangan data, yang mempengaruhi kemampuan model dalam mengenali kasus strok secara merata. Penelitian selanjutnya dianjurkan untuk menerapkan teknik penyeimbangan data serta menggabungkan Random Forest dengan algoritma lain guna meningkatkan akurasi dan keandalan prediksi.

Kata Kunci : Strok, *Random Forest*, *Knowledge Discovery in Database*, Prediksi

Abstract - Stroke is a condition when blood flow to the brain is disrupted, which can lead to brain damage and even death. Stroke is also the leading cause of death and disability worldwide. With a high accident rate, early detection of stroke is essential to reduce the risk of stroke. This study aims to classify stroke risk based on the Random Forest algorithm through the Knowledge Discovery in Database (KDD) approach. The dataset used is a dataset from the Kaggle platform consisting of 35,000 patient data with 18 relevant attributes that have gone through a cleaning process and selection features to achieve optimal prediction accuracy. The training model is carried out by dividing data by dividing 80% of training data and 20% of testing data. The results of the evaluation using the confusion matrix showed that the Random Forest Algorithm had an accuracy of 87.47% with a precision in class 1 of 96.40% and a recall of 68.53%, and a precision in class 0 of 84.51% and a recall of 98.51%. However, the main challenge in this study is the imbalance of the data, which affects the model's ability to recognize stroke cases evenly. Further research is recommended to apply data balancing techniques and combine Random Forest with other algorithms to improve the accuracy and reliability of predictions.

Keywords: Stroke, *Random Forest*, *Knowledge Discovery in Database*, Prediction

I. PENDAHULUAN

Strok adalah kondisi ketika aliran darah ke otak terganggu yang disebabkan oleh berkurangnya pasokan darah. Hal ini dapat mengakibatkan kecacatan hingga kematian (Dewi & Fitraneti, 2024). Menurut data dari organisasi strok dunia atau *World Stroke Organization* strok menempati peringkat ke-2 sebagai penyebab kematian di seluruh dunia dengan sekitar 7 juta kematian, dan 12 juta kasus baru setiap tahunnya. Sekitar 1 dari 4 orang diperkirakan

akan mengalami strok dalam hidupnya dan 53% terjadinya strok pada seseorang berusia dibawah 70 tahun (Feigin et al., 2025). Data dari survei Kesehatan Indonesia pada tahun 2023 prevalansi dari kejadian strok di Indonesia sebanyak 8,3 per 1.000 penduduk dan sebesar 11,2% dari total kecacatan dan 18,5% dari total kematian yang terjadi pada Indonesia penyebab utamanya adalah karena strok (Kemenkes, 2024).

Semakin bertambahnya usia semakin tinggi kemungkinan untuk terkena stroke (Hisni et al., 2022). Selain usia yang menjadi penyebab terkena stroke terdapat juga faktor-faktor lain yang membuat terjadinya stroke yaitu dikarenakan riwayat keluarga yang pernah mengalami stroke atau penyakit jantung, kolesterol tinggi, diabetes, merokok, dan konsumsi alkohol (Widiani & Yasa, 2023). berdasarkan data-data yang telah dikaji bahwa stroke merupakan masalah yang sangat serius terhadap kesehatan manusia dan dapat menyebabkan kelumpuhan atau bahkan kematian.

Perkembangan teknologi saat ini terkhususnya pada bidang kecerdasan buatan dan *machine learning* memungkinkan dalam mendeteksi penyakit berdasarkan dari data medis. Salah satu algoritma yang banyak digunakan adalah Algoritma *Random Forest* yang mampu mengklasifikasikan data berdasarkan sejumlah factor Kesehatan yang berperan dalam risiko stroke.

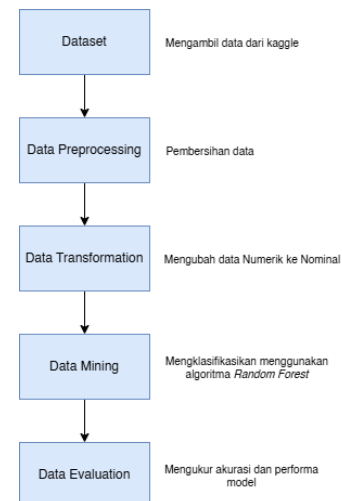
Penelitian sebelumnya sudah banyak membahas tentang prediksi risiko stroke menggunakan berbagai algoritma machine learning. Misalnya, pada penelitian yang dilakukan oleh Murdiansyah yang membahas prediksi stroke menggunakan *Extreme Gradient Boosting* (XGBoost), dalam penelitian ini menemukan bahwa metode ini memiliki akurasi cukup tinggi, dengan tingkat akurasi mencapai 92,5% dalam memprediksi stroke berdasarkan faktor yang sudah ditentukan (Murdiansyah, 2024). Penelitian lain membandingkan algoritma *Decision Tree*, *Naive Bayes*, dan *Support Vector Machine* (SVM) dalam klasifikasi risiko stroke, disini algoritma *Decision Tree* menunjukkan akurasi sebesar (94%), SVM (89%) dan *Naive Bayes* (85%) (Aulia et al., 2024). Selain itu penelitian dari Maskuri memprediksi stroke menggunakan algoritma *K-Nearest Neighbor* (KNN), penelitian ini menggunakan metode split validation dengan 80% rasio data latih dan 20% data diuji untuk mengevaluasi kinerja model, dengan parameter $k = 9$ model memiliki tingkat akurasi 95% dalam mengklasifikasi pasien dengan memiliki diagnosa penyakit stroke (Maskuri et al., 2022).

Penelitian yang dilakukan oleh Azhar menunjukkan bahwa algoritma *Random Forest* memiliki tingkat akurasi yang sangat tinggi dalam memprediksi berbagai penyakit dengan data medis, hasil akurasi yang dilakukan mencapai 96%, hasil ini juga menjadi tingkat akurasi paling tinggi dibandingkan beberapa algoritma lainnya (Azhar et al., 2022).

Berdasarkan dari semua penelitian tersebut, sangat terlihat bahwa setiap algoritma memiliki keunggulan dan keterbatasan masing-masing dalam memprediksi penyakit stroke, semuanya tergantung pada parameter yang digunakan serta karakteristik dataset. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasi risiko stroke menggunakan algoritma *Random Forest*. Kemampuan metode ini dalam menangani data yang kompleks dan menghasilkan akurasi yang tinggi membuat metode ini menjadi pilihan untuk digunakan.

II. METODOLOGI PENELITIAN

Proses penelitian dilakukan dengan beberapa tahapan utama berdasarkan pendekatan *Knowledge Discovery in Database* (KDD). yaitu pengumpulan data, pemrosesan data, transformasi data, penerapan algoritma, dan evaluasi model.



Sumber: Peneliti (2025)

Gambar 1. Diagram Alur Proses Klasifikasi Risiko Stroke dengan pendekatan KDD.

1. Teknik Pengumpulan Data

Teknik pengumpulan data yang dpergunakan dalam penelitian ini adalah bersumber dari Kumpulan dataset yang ada di situs *Kaggle* dengan link <https://www.kaggle.com/datasets/mahatiratusher/stroke-risk-prediction-dataset-v2>. Dataset yang diperoleh dengan cara mengcopy table data dan menyimpannya menjadi tabel dataset dengan jumlah 35.000 orang yang memiliki gejala penyakit stroke seperti yang ditampilkan pada Tabel 1.

Tabel 1. *Dataset* Risiko stroke

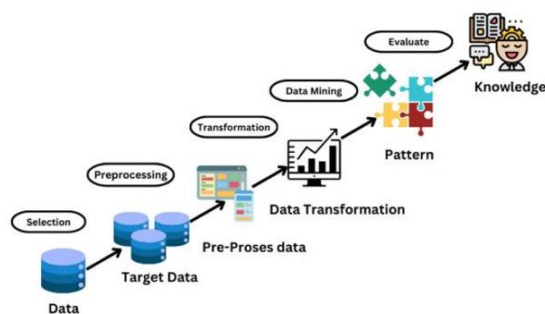
Atribut	Keterangan
---------	------------

Age	Umur/usia
Gender	Jenis Kelamin (Male/Female)
Chest_pain	Ada gejala (1) / tidak ada gejala (0)
Shortness_of_breath	Ada gejala (1) / tidak ada gejala (0)
Irregular_heartbeat	Ada gejala (1) / tidak ada gejala (0)
Fatigue_weakness	Ada gejala (1) / tidak ada gejala (0)
Dizziness	Ada gejala (1) / tidak ada gejala (0)
Swelling_edema	Ada gejala (1) / tidak ada gejala (0)
Neck_Jaw_Pain	Ada gejala (1) / tidak ada gejala (0)
Excessive_Sweating	Ada gejala (1) / tidak ada gejala (0)
Persistent_Cough	Ada gejala (1) / tidak ada gejala (0)
Nausea_vomiting	Ada gejala (1) / tidak ada gejala (0)
High_blood_pressure	Ada gejala (1) / tidak ada gejala (0)
Chest_discomfort	Ada gejala (1) / tidak ada gejala (0)
Cold_hands_feet	Ada gejala (1) / tidak ada gejala (0)
Snoring_sleep_apnea	Ada gejala (1) / tidak ada gejala (0)
Anxiety_disorder	Ada gejala (1) / tidak ada gejala (0)
at_risk	beresiko (1) / tidak beresiko (0)
stroke_risk_percentage	persentase risiko stroke (0-100%)

Sumber: Kaggle (2025)

2. Tahapan Knowledge Discovery in Database (KDD)

Metode Pengolahan data yang digunakan pada penelitian ini adalah metode *Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* (KDD) merupakan tahapan sistematis yang dilakukan untuk mendapatkan pengetahuan baru dari kumpulan data yang besar dan kompleks (Rois et al., 2025).



Sumber: (Buani, 2024)

Gambar 2. Model Proses *Knowledge Discovery in Database* (KDD).

Pada Gambar 1. menggambarkan tentang tahapan-tahapan pada *Knowledge Discovery in Database* yang mana terdiri dari tahapan *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Data Mining* dan *Evaluation*.

a. Data Selection

Data Selection adalah tahapan dalam pengambilan dan pemilihan data dari sebuah *dataset* yang nantinya akan diolah (Alghifari & Juardi, 2021).

Pada tahapan ini akan dipilih atribut-atribut yang relevan dengan risiko stroke berdasarkan *dataset* asli. Seleksi ini dilakukan untuk memastikan hanya fitur yang berkontribusi terhadap klasifikasi yang digunakan

b. Data Preprocessing

Data Preprocessing adalah tahapan awal dalam pengolahan data awal yang digunakan untuk mengubah data mentah yang diperoleh dari berbagai sumber agar menjadi informasi yang lebih bersih dan siap digunakan untuk analisis lebih lanjut (Salam et al., 2023).

Pada tahap ini data yang telah dipilih kemudian diperiksa untuk memastikan bahwa tidak ada nilai kosong (*Missing Value*) maupun duplikasi data.

c. Data Transformation

Data Transformation adalah tahapan untuk mengubah data menjadi format yang lebih sesuai agar dapat diolah pada proses selanjutnya (Supoyo & Prasetyaninrum, 2022).

Pada tahap ini data ditransformasi gunanya untuk menyesuaikan format dari atribut dengan kebutuhan dari algoritma yang ingin dipilih

d. Data Mining

Data Mining adalah tahapan untuk menggali informasi yang berguna dari suatu data besar baik itu terstruktur maupun tidak terstruktur dengan menggunakan metode atau algoritma tertentu (Fitria & Rozci, 2022). Pada tahap ini dilakukan Pembangunan model klasifikasi menggunakan algoritma *Random Forest*. *Random Forest* merupakan salah satu metode atau algoritma dari *machine learning* yang terdiri dari gabungan beberapa *Decision Tree* yang mana digunakan untuk mengklasifikasi data ke suatu kelas (Suci Amaliah et al., 2022). *Random Forest* dipilih karena kemampuannya dalam menangani data berjumlah besar, memiliki banyak atribut, dan

kemampuannya untuk mengurangi overfitting melalui agregasi dari banyak decision tree.

e. **Evaluation**

Evaluation adalah tahapan untuk mendapatkan pola informasi yang dihasilkan dari tahapan *data mining* yang ditampilkan ke dalam bentuk yang mudah dipahami (Naldy & Andri, 2021).

Model yang telah dibangun kemudian dievaluasi menggunakan confusion matrix untuk mengukur akurasi dari model, kemudian nilai precision dan recall pada masing-masing kelas. *Confusion Matrix* merupakan metode yang digunakan untuk mengukur kinerja dari model klasifikasi khususnya dalam bidang data mining (Romadloni et al., 2022).

Tabel 2. *Confusion Matrix*

Predicted	Actual	
	True	False
True	TP (True Positive)	FP (False Positive)
False	TN (True Negative)	FN (False Negative)

Sumber: (Romadloni et al., 2022)

Pada tabel tersebut, terdapat empat istilah yang mewakili hasil dari proses klasifikasi confusion matrix, yaitu:

- 1) *True Positive* (TP), data positif yang berhasil diprediksi dengan benar.
- 2) *True Negative* (TN), data negatif yang berhasil diprediksi dengan benar.
- 3) *False Positive* (FP), data negatif yang berhasil diprediksi dengan benar.
- 4) *False Negative* (FN), data negatif yang berhasil diprediksi dengan negatif.

Berdasarkan dari confusion matrix tersebut dapat dihitung beberapa nilai Accuracy, Precision dan Recall.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

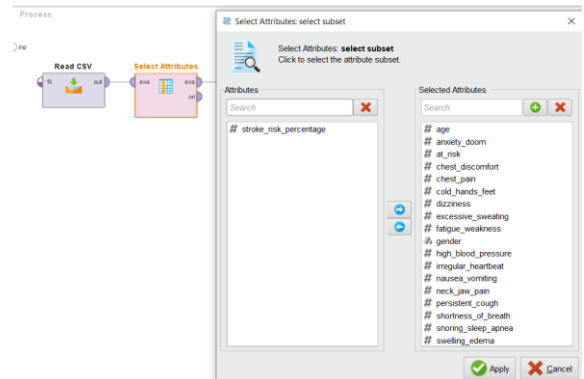
III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset yang berjumlah 35.000 data pasien dengan 18 atribut yang diambil dari website dataset Kaggle

1. Data Selection

Data atau atribut diambil dari *dataset* risiko strok di situs *kaggle* yang berjumlah 18 atribut

dari 19 atribut, dimana atribut yang tidak diambil adalah atribut dari *strok_risk_percentage*.



Sumber: Peneliti (2025)

Gambar 3. *Data Selection* pada *Rapidminer*

2. Data Preprocessing

Data yang telah dipilih sebelumnya dilakukan proses pengecekan terhadap keberadaan data yang hilang (*missing value*). Hasil dari pengecekan tersebut menunjukkan bahwa tidak terdapat data yang hilang pada seluruh atribut dalam dataset. Hal ini menunjukkan bahwa dataset memiliki kualitas yang baik dari segi kelengkapan data, sehingga proses analisis dapat dilanjutkan tanpa perlu melakukan teknik imputasi atau pengisian data kosong. Gambar 4. berikut menampilkan hasil pengecekan *missing value* yang dilakukan menggunakan aplikasi *RapidMiner*.

Name	Type	Missing	Statistics	Filter (18 / 18 attributes)	Search for Attributes
age	Integer	0	Min: 18, Max: 86, Average: 38.630		
gender	Nominal	0	Count: Male (17472), Female (17528), Female (17528), Male (17472)		
chest_pain	Integer	0	Min: 0, Max: 1, Average: 0.148		
high_blood_pressure	Integer	0	Min: 0, Max: 1, Average: 0.252		
irregular_heartbeat	Integer	0	Min: 0, Max: 1, Average: 0.098		
shortness_of_breath	Integer	0	Min: 0, Max: 1, Average: 0.198		
fatigue_weakness	Integer	0	Min: 0, Max: 1, Average: 0.245		
dizziness	Integer	0	Min: 0, Max: 1, Average: 0.191		
swelling_edema	Integer	0	Min: 0, Max: 1, Average: 0.148		

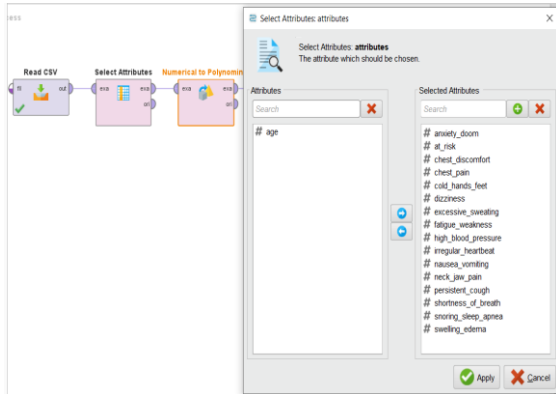
Sumber: Peneliti (2025)

Gambar 4. Pengecekan *Missing Value* pada *Rapidminer*

3. Data Transformation

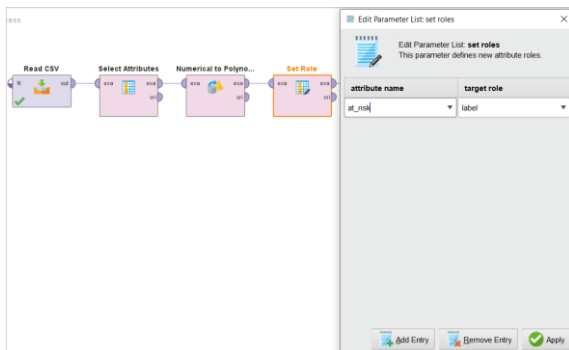
Data yang sudah dibersihkan diubah menjadi format yang lebih sesuai agar pada tahapan selanjutnya data tersebut dapat diolah. Pada penelitian ini data yang diubah adalah data dari *anxiety_doom*, *at_risk*, *chest_discomfort*, *chest_pain*, *cold_hands_feet*, *dizziness*, *excessive_sweating*, *fatigue_weakness*, *high_blood_pressure*, *irregular_heartbeat*, *nausea_vomiting*, *neck_jaw_pain*, *persistent_cough*,

shortness_of_breath, snoring_sleep_apnea, dan swelling_edema yang mana data tersebut diolah dari data numerical menjadi data nominal agar dapat diolah pada proses selanjutnya.



Sumber: Peneliti (2025)
 Gambar 5. Pengubahan Data dari Numerical menjadi Nominal pada Rapidminer.

Setelah datanya diubah menjadi nominal selanjutnya data akan diubah role attribute at_risk yang akan dijadikan sebagai label.

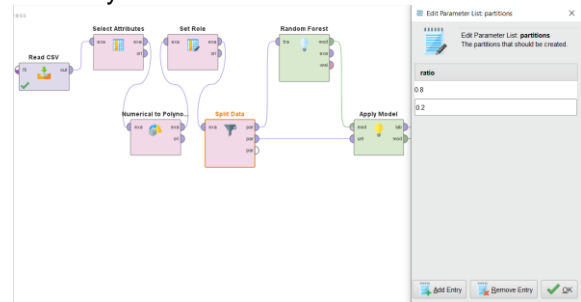


Sumber: Peneliti (2025)
 Gambar 6. Perubahan Role attribute at_risk menjadi label pada Rapidminer

4. Data Mining

Data yang telah melalui proses pembersihan dan transformasi selanjutnya dibagi menjadi dua bagian, yaitu data training dan data testing. Pembagian dilakukan dengan rasio 80% untuk data training dan 20% untuk data testing. Data training digunakan untuk membangun dan melatih model klasifikasi menggunakan algoritma Random Forest, sedangkan data testing digunakan untuk menguji kinerja dan mengukur tingkat keakuratan prediksi dari model yang telah dibangun. Pembagian data ini bertujuan agar model dapat dievaluasi secara objektif terhadap data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi lebih merepresentasikan performa model dalam

kasus nyata.



Sumber: Peneliti (2025)
 Gambar 7. Pembagian Data Training dan Testing

4. Data Evaluation

Data yang telah diklasifikasikan menggunakan algoritma Random Forest selanjutnya dievaluasi untuk mengukur performa model dalam melakukan prediksi. Evaluasi dilakukan dengan menggunakan metode confusion matrix, yang berfungsi untuk menilai sejauh mana model mampu membedakan antara kelas yang berisiko dan tidak berisiko terhadap stroke. Confusion matrix menyajikan hasil klasifikasi dalam bentuk matriks yang menunjukkan jumlah prediksi benar dan salah untuk masing-masing kelas. Proses evaluasi ini dilakukan melalui aplikasi RapidMiner, dan hasil visualisasi dari confusion matrix ditampilkan pada

Gambar 8.

accuracy: 87.47%			
	true 0	true 1	class precision
pred 0	4357	911	84.31%
pred 1	66	1766	96.40%
class recall	99.51%	68.53%	

Sumber: Peneliti (2025)
 Gambar 8. Confusion Matrix Algoritma Random Forest

Berikut adalah hasil perhitungan manual Confusion Matrix.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{1766 + 4357}{1766 + 4357 + 66 + 811} = \frac{6123}{7000} = 0,8747$$

$$Precision = \frac{TP}{TP + FP} = \frac{1766}{1766 + 66} = \frac{1766}{1832} = 0,9639$$

$$Recall = \frac{TP}{TP + FN} = \frac{1766}{1766 + 811} = \frac{1766}{2577} = 0,6853$$

Sumber: Peneliti (2025)
 Gambar 9. Perhitungan Manual Confusion Matrix

Tabel 3. Hasil Perhitungan Confusion Matrix

	Nilai(%)
Accuracy	87.47%
Precision	84.31%
Recall	68.53%

Sumber: Peneliti (2025)

Hasil *confusion matrix* menunjukkan akurasi sebanyak 87.47%. Hasil klasifikasi model mendapatkan sampel sebanyak 4357 dari kelas 0 dan 1766 sampel dari kelas 1. *Precision model* mendapatkan 84.31% untuk kelas 0, dan 96.40% untuk kelas 1. *Recall* tertinggi didapatkan dari kelas 0 sebanyak 98.51%, dan *recall* terendah sebanyak 68.53% yang didapatkan dari kelas 1. Hasil ini menunjukkan bahwa model cenderung lebih sensitif dalam mengidentifikasi individu yang tidak berisiko dibandingkan dengan individu yang berisiko. Tingginya recall pada kelas 0 menunjukkan bahwa hampir semua kasus tidak berisiko berhasil dikenali. Sebaliknya, recall yang lebih rendah pada kelas 1 mengindikasikan bahwa model masih mengalami kesulitan dalam mendeteksi seluruh kasus yang benar-benar berisiko.

Tabel 4. Hasil Penyebab Faktor terhadap Model

Faktor	Pengaruh terhadap Model
Ketidakeimbangan Kelas	Model lebih fokus mengenali kelas mayoritas (tidak berisiko)
Karakteristik Fitur	Tidak semua fitur relevan terhadap risiko stroke
Parameter Model <i>Random Forest</i>	Setting parameter kurang optimal bisa menyebabkan overfitting/underfitting

Sumber: Peneliti (2025)

Berdasarkan hasil tersebut, dapat disimpulkan bahwa meskipun model memiliki akurasi tinggi masih terdapat hal ketidakeimbangan dalam mengenali kelas 1 dengan baik. Hal ini dapat berdampak pada kinerja model dalam kasus nyata yang membutuhkan identifikasi kelas 1 secara akurat.

IV. KESIMPULAN

Penelitian menggunakan algoritma *Random Forest* telah berhasil diimplementasi dalam mengklasifikasikan risiko stroke, melalui pendekatan *Knowledge Discovery In Database* (KDD) metode ini memiliki akurasi yang tinggi. Pada *confusion matrix* akurasi model sebesar 87,47% dengan 4357 sampel dari kelas 0 dan 1766 sampel dari kelas 1. Pada *precision model* mencapai 84,31% untuk kelas 0 dan 96,40% untuk kelas 1, sementara itu recall terendah ada pada kelas 1 sebesar 68,53%, sedangkan pada kelas 0 itu recall tertinggi sebesar 98,51%. Hasil ini menunjukkan meskipun model memiliki akurasi yang tinggi tetapi masih terdapat

ketidakeimbangan dalam mengenali kelas 1 dengan baik, yang bisa berdampak pada kinerja model dalam aplikasi nyata. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan balancing data, memanfaatkan data real time guna meningkatkan keandalan model dalam aplikasi medis.

Pengembangan lebih lanjut dapat difokuskan pada analisis faktor risiko spesifik, seperti pola makan, tingkat stres, dan gaya hidup, serta pemanfaatan teknologi dalam pemantauan kesehatan. Studi lanjutan juga diharapkan dapat mengintegrasikan metode prediksi berbasis big data dan analisis genetik untuk memahami lebih lanjut predisposisi seseorang terhadap stroke. Selain itu, penelitian dengan cakupan populasi lebih luas dan pendekatan interdisipliner dapat memberikan rekomendasi lebih aplikatif dalam upaya pencegahan serta pengelolaan risiko stroke secara efektif.

V. REFERENSI

- Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika (JIF)*, 9(2).
- Aulia, Y., Andriyansyah, A., Suharjo, S., & Nensi, S. W. (2024). Analisis Prediksi Stroke dengan Membandingkan Tiga Metode Klasifikasi Decision Tree, Naïve Bayes, dan Random Forest. *Jurnal Ilmu Komputer Dan Informatika*, 3(2), 89–98. <https://doi.org/10.54082/jiki.90>
- Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. *Science and Information Technology SINTECH Journal*, 5(2), 191–197.
- Buani, D. C. P. (2024). Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest. *Evolusi: Jurnal Sains Dan Manajemen*, 12(1).
- Dewi, L., & Fitraneti, E. (2024). Stroke Iskemik. *Scientific Journal*, 3(6), 379.
- Feigin, V. L., Brainin, M., Norrving, B., Martins, S. O., Pandian, J., Lindsay, P., F Grupper, M., & Rautalin, I. (2025). World Stroke Organization: Global Stroke Fact Sheet 2025. *International Journal of Stroke*, 20(2), 132–144. <https://doi.org/10.1177/17474930241308142>
- Fitria, E. R., & Rozci, F. (2022). PENERAPAN METODE REGRESI LEAST ABSOLUTE SHRINKAGE AND SELECTION

- OPERATOR (LASSO) DAN REGRESI LINIER UNTUK MEMPREDIKSI TINGKAT KEMISKINAN DI INDONESIA. *Jurnal Ilmiah Sosio Agribis (JISA)*, 22(2), 123–132.
- Hisni, D., Saputri, M. E., & Sujarni. (2022). Faktor - faktor yang Berhubungan dengan Kejadian Stroke Iskemik di Instalasi Fisioterapi Rumah Sakit Pluit Jakarta Utara Periode Tahun 2021. *Jurnal Penelitian Keperawatan Kontemporer*, 2(1), 140–149.
- Kemkes. (2024, October 25). *Cegah Stroke dengan Aktivitas Fisik*. Kementerian Kesehatan Republik Indonesia. <https://kemkes.go.id/id/rilis-kesehatan/cegah-stroke-dengan-aktivitas-fisik>
- Maskuri, M. N., Harliana, Sukerti, K., & Bhakti, R. M. H. (2022). Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 4(1), 130–140.
- Murdiansyah, D. T. (2024). Prediksi Stroke Menggunakan Extreme Gradient Boosting. *JIKO (Jurnal Informatika Dan Komputer)*, 8(2), 419–426.
- Naldy, E. T., & Andri. (2021). Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN. *Jurnal Nasional Ilmu Komputer*, 2(2), 89–101.
- Rois, Moh. I. A., Dwilestari, G., & Suarna, N. (2025). PREDIKSI PERSETUJUAN PINJAMAN MENGGUNAKAN DATASET LOAN APPROVAL MENGGUNAKAN ALGORITMA KLASIFIKASI. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 1342–1347.
- Romadloni, P. L., Kusuma, B. A., & Baihaqi, W. M. (2022). Komparasi Metode Pembelajaran Mesin untuk Implementasi Pengambilan Keputusan dalam Menentukan Promosi Jabatan Karyawan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 622–628.
- Salam, R. R., Jamil, M. F., Ibrahim, Y., Rahmadden, Soni, & Herianto. (2023). Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 27–35.
- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127. <https://doi.org/10.35580/variasiunm31>
- Supoyo, A., & Prasetyaninrum, P. T. (2022). Analisis Data Mining Untuk Memprediksi Lama Perawatan Pasien Covid-19 Di DIY. *Biaglala Infromatika*, 10(1), 21–29.
- Widiani, G. A. R., & Yasa, I. M. M. (2023). Korelasi Tingkat Pengetahuan Terhadap Kemampuan Deteksi Dini Gejala Stroke dengan Sikap Keluarga Terhadap Penanganan Pre Hospital. *Bina Generasi: Jurnal Kesehatan*, 2(14), 25–30.