

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**PERBANDINGAN METODE PENGUKURAN JARAK PADA
KINERJA ALGORITMA *K-NEAREST NEIGHBOR* (KNN)
DALAM KLASIFIKASI PENYAKIT *STROKE***

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

SHAFIQ ZAHIDAN

NIM. 12050110329



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2025**



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Shafiq Zahidan
 NIM : 12050110329
 Tempat/Tgl.Lahir : Pekanbaru/4 Juli 2001
 Fakultas : Sains dan Teknologi
 Prodi : Teknik Informatika
 Judul Skripsi : PERBANDINGAN METODE PENGUKURAN JARAK PADA KINERJA ALGORITMA K-NEAREST NEIGHBOR (KNN) DALAM KLASIFIKASI PENYAKIT STROKE

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikian surat pernyataan ini saya buat dengan penuh kesadaran dan akal sehat.

Pekanbaru, 7 Januari 2025
 Yang membuat pernyataan



Shafiq Zahidan
NIM. 12050110329



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

PERBANDINGAN METODE PENGUKURAN JARAK PADA KINERJA ALGORITMA *K-NEAREST NEIGHBOR* (KNN) DALAM KLASIFIKASI PENYAKIT *STROKE*

TUGAS AKHIR

Oleh

SHAFIQ ZAHIDAN

NIM. 12050110329

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir di Pekanbaru, pada tanggal 7 Januari 2025

Pembimbing,

Siska Kurnia Gusti, S.T., M.Sc.

NIP. 198610092022032001

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

PERBANDINGAN METODE PENGUKURAN JARAK PADA KINERJA ALGORITMA *K-NEAREST NEIGHBOR* (KNN) DALAM KLASIFIKASI PENYAKIT *STROKE*

Oleh

SHAFIQ ZAHIDAN

NIM. 12050110329

Telah dipertahankan di depan sidang dewan penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Universitas Islam Negeri Sultan Syarif Kasim Riau


Pekanbaru, 7 Januari 2025

Mengesahkan,

Ketua Jurusan,

Dekan,

DR. HARTONO, M.Pd.
NIP. 19640301 199203 1 003


IWAN ISKANDAR, M.T.
NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua : Prof. Okfalisa, S.T., M.Sc.
Pembimbing I : Siska Kurnia Gusti, S.T., M.Sc.
Penguji I : Fadhilah Syafria, S.T., M.Kom
Penguji II : Iis Afrianty, S.T., M.Sc.



UIN SUSKA RIAU



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 7 Januari 2025

Yang membuat pernyataan,

SHAFIQ ZAHIDAN

12050110329

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah Rabbil'alamiin...

Rasa syukur kuhaturkan kepada-Mu, Yaa Allah yang Maha Ber-Ilmu, hanya karena karunia-Mu lah hamba-Mu akhirnya dapat menyelesaikan Tugas Akhir ini

Allah Azza Wa Jalla

Sholawat serta salam untuk Rasulullah

Muhammad Shallallahu 'Alaihi Wassalam

Kupersembahkan karya sederhana ini, Tugas Akhir ini, untuk Ibu dan Ayah. Tentulah tidak akan tergantikan semua jasa, pengorbanan, tetes keringat dan rasa letih itu, hanya dengan karya kecil dariku ini. Namun semoga dengan ini, aku dapat mengukir sebaris senyum bahagia di hati Ibu dan Ayah. Jika boleh kujabarkan cinta, tentulah tidak pernah dapat seindah rasa syukurku menjadi anakmu. Terimakasih untuk semua rangkaian do'a, kasih sayang serta ilmu yang berharga.

Dan tidak lupa kupersembahkan untuk semua saudara kandung ku yang tersayang, terimakasih untuk semua do'a dan dukungan yang telah diberikan selama ini.

Juga, kupersembahkan untuk semua keluargaku dan kerabat. Semua kesulitan seolah lenyap saat mengingat bahwa aku memiliki dukungan dan do'a darimu semua. Aku tahu, engkau semua berjuang lebih keras dariku, namun selalu memiliki energi yang hebat untuk menyemangatiku.

Alhamdulillah, Allah menganugerahiku keluarga, kerabat dan kehidupan yang indah

ABSTRAK

Stroke adalah kondisi di mana fungsi saraf terganggu karena gumpalan darah yang mengalami penyumbatan. *Stroke* adalah salah satu penyakit ketiga paling umum di dunia yang menyebabkan kecacatan. Menurut data yang dikumpulkan oleh Badan Riset Kesehatan Indonesia, *Stroke* menyumbang 21,1% kematian di Indonesia. Penelitian ini menggunakan Metode *K-Nearest Neighbor* untuk klasifikasi penyakit *stroke* dengan membandingkan perhitungan jarak (*Euclidean, Manhattan, Minkowski, Chebyshev dan Cosine*). Data menggunakan *Kaggle Dataset* yang terdiri dari 15000 data. Dengan pembagian rasio data latih dan data uji 90:10, 80:20 dan 70:30. Penelitian ini menggunakan beberapa pengujian yaitu pengujian menggunakan fitur yang telah diseleksi oleh dokter spesialis saraf, pengujian menggunakan semua fitur pada dataset, pengujian menggunakan *feature selection* pada dataset dan pengujian menggunakan rapid miner. Hasil pengujian Akurasi tertinggi menggunakan pengujian semua fitur pada dataset penyakit *stroke*, dengan nilai akurasi 53.33%, presisi 53.33%, recall 53.33%, rasio perbandingan 90:10 menggunakan perhitungan jarak euclidean.

Kata Kunci: Data Mining, Klasifikasi, K-Nearest Neighbor, Perhitungan Jarak, Penyakit *Stroke*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



ABSTRACT

Stroke is a condition in which nerve function is disrupted due to a blood clot that is blocked. Stroke is one of the third most common diseases in the world that causes disability. According to data collected by the Indonesian Health Research Agency, Stroke accounts for 21.1% of deaths in Indonesia. This study uses the K-Nearest Neighbor Method for the classification of stroke disease by comparing distance calculations (Euclidean, Manhattan, Minkowski, Chebyshev and Cosine). Data using the Kaggle Dataset consisting of 15000 data. With the division of the training data ratio and test data 90:10, 80:20 and 70:30. This study uses several tests, namely testing using features that have been selected by a neurologist, testing using all features in the dataset, testing using feature selection in the dataset and testing using rapid miner. The highest accuracy test results use testing all features in the stroke disease dataset, with an accuracy value of 53.33%, precision 53.33%, recall 53.33%, comparison ratio 90:10 using euclidean distance calculations.

Keywords: Classification, Data Mining, K-Nearest Neighbor Distance Calculation, Stroke Disease.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalammu 'alaikum wa rohmatullohi wa barokatuh.

Alhamdulillah *robbil'alamin*, tak henti-hentinya kami ucapkan kehadiran Allah *Subhanahu wa ta'ala*, yang dengan rahmat dan hidayah-Nya kami mampu menyelesaikan Tugas Akhir ini dengan baik. Tidak lupa bershalawat kepada Nabi dan Rasul-Nya, Nabi Muhammad *Sholallohu 'alaihi wa salam*, yang telah membimbing kita sebagai umatnya menuju jalan kebaikan.

Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar sarjana pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Banyak sekali pihak yang telah membantu kami dalam penyusunan laporan ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada kami. Semua itu tentu terlalu banyak bagi kami untuk membalasnya, namun pada kesempatan ini kami hanya dapat mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Khairunnas Rajab, M.Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Iwan Iskandar, M.T. selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Bapak Dr Alwis Nazir, M.Kom., selaku Pembimbing Akademik yang telah memberikan bimbingan, motivasi, semangat dan ilmu yang bermanfaat.
5. Ibu Siska Kurnia Gusti, S.T., M.Sc., selaku pembimbing Tugas Akhir. Terimakasih sudah meluangkan waktu untuk membimbing serta memberi arahan dan motivasi sehingga laporan tugas akhir ini dapat terselesaikan dengan baik.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6. Ibu Fadhilah Syafria, S.T., M.Kom., selaku Penguji I dan Ibu Iis Afrianty, S.T., M.Sc., selaku penguji II Tugas Akhir yang sudah memberikan kritik, saran serta arahan dalam menyelesaikan tugas akhir ini.
7. Bapak/Ibu dosen program studi Teknik Informatika yang telah memberikan ilmu kepada penulis selama masa perkuliahan.
8. Orang tua tersayang, Ibu Darma Yunita dan Bapak Zamharil yang selalu memberi dukungan, semangat dan mendo'akan tanpa henti hingga selesainya tugas akhir ini.
9. Ke-dua Adik Kandungku yang selalu menghibur dan menyemangati penulis dalam menyelesaikan laporan tugas akhir ini.
10. Skuy teman seperjuangan dari Maba yang selalu ada dan sangat membantu penulis dalam menyelesaikan Tugas Akhir ini.
11. Semua teman-teman yang sudah membantu, menyemangati, dan menghibur sehingga penulis dapat menyelesaikan laporan tugas Akhir ini.

Demikian tugas akhir ini dibuat, semoga dapat bermanfaat khususnya untuk penulis. Kami menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat kami harapkan untuk kesempurnaan laporan ini dapat disampaikan melalui email 12050110329@students.uin-suska.ac.id. Akhirnya kami berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya.

Wassalamu 'alaikum wa rohmatullohi wa barokatuh.

Pekanbaru, 7 Januari 2025

Penulis

DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN.....	Error! Bookmark not defined.
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN.....	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
DAFTAR RUMUS	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB 2 KAJIAN PUSTAKA.....	5
2.1 Data Mining.....	5
2.1.1 <i>Knowledge Discovery in Database</i>	5
2.2 Klasifikasi	7
2.3 <i>K-nearest neighbor (KNN)</i>	8

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.4	<i>Confusion Matrix</i>	10
2.5	<i>Stroke</i>	12
2.5.1	Faktor Risiko <i>Stroke</i>	13
2.6	Penelitian Terkait.....	15
BAB 3 METODOLOGI PENELITIAN		22
3.1	Tahapan Penelitian.....	22
3.2	Perumusan Masalah	23
3.3	Studi Literatur.....	23
3.4	Pengumpulan Data.....	23
3.5	Analisa Data	24
3.5.1	Normalisasi Data	25
3.6	<i>K-nearest neighbor</i>	26
3.7	Implemetasi dan Pengujian	27
3.7.1	Implementasi	27
3.7.2	Pengujian.....	28
3.7.3	Evaluasi.....	28
3.8	Kesimpulan dan Saran	29
BAB 4 PEMBAHASAN		30
4.1	Analisa Kebutuhan Data	30
4.1.1	Pengumpulan Data.....	30
4.2	Analisa Proses KDD	31
4.2.1	Data Selection.....	31
4.2.2	Pre-processing	33
4.2.3	Transformasi Data	34
4.2.4	Klasifikasi <i>K-nearest neighbor</i>	37



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.3	Implementasi	55
4.3.1	Batasan Implementasi	56
4.3.2	Lingkungan Implementasi.....	56
4.3.3	Langkah Implementasi.....	56
4.4	Evaluasi.....	59
4.1.1	Hasil Skenario Pengujian	61
4.1.2	Analisa Pengujian.....	63
BAB 5 PENUTUP		81
5.1	Kesimpulan.....	81
5.2	Saran	81
DAFTAR PUSTAKA		82
DAFTAR RIWAYAT HIDUP.....		85

DAFTAR GAMBAR

Gambar 1 Tahapan KDD.....	6
Gambar 2 Tahapan Metodologi Penelitian.....	22
Gambar 3 <i>Flowchart</i> Proses Pelatihan.....	26
Gambar 4 Mengimport Dataset	56
Gambar 5 Model KNN dan Mengevaluasi Hasil.....	57
Gambar 6 Evaluasi Metric Jarak dan K	58
Gambar 7 Tampilan Confusion Matrix 90:10	60
Gambar 8 Tampilan Confusion Matrix 80:20	60
Gambar 9 Tampilan Confusion Matrix 70:30	60
Gambar 10 Tampilan Hasil Classification Report.....	61
Gambar 11 <i>Confusion Matrix</i> Semua Fitur 90:10	63
Gambar 12 <i>Confusion Matrix</i> Semua Fitur 80:20	64
Gambar 13 <i>Confusion Matrix</i> Semua Fitur 70:30	64
Gambar 14 <i>Confusion Matrix Feature Selection</i> 90:10.....	66
Gambar 15 <i>Confusion Matrix Feature Selection</i> 80:20.....	67
Gambar 16 <i>Confusion Matrix Feature Selection</i> 70:30.....	67
Gambar 17 Akurasi <i>Rapid Miner</i> 90:10.....	69
Gambar 18 Akurasi <i>Rapid Miner</i> 80:20.....	70
Gambar 19 Akurasi <i>Rapid Miner</i> 70:30.....	70
Gambar 20 Visualisasi Age dengan Diagnosis	71
Gambar 21 Visualisasi Gender dengan Diagnosis	71
Gambar 22 Visualisasi <i>Hypertension</i> dengan Diagnosis	72
Gambar 23 Visualisasi <i>Heart Disease</i> dengan Diagnosis.....	72
Gambar 24 Visualisasi <i>Average Glucose Level</i> dengan Diagnosis	73
Gambar 25 Visualisasi <i>Body Mass Index</i> (BMI) dengan Diagnosis.....	74
Gambar 27 Visualisasi <i>Smoking Status</i> dengan Diagnosis	75
Gambar 28 Visualisasi <i>Alcohol Intake</i> dengan Diagnosis	75
Gambar 29 Visualisasi <i>Dietary Habits</i> dengan Diagnosis	76
Gambar 30 Visualisasi HDL dengan Diagnosis	77
Gambar 31 Visualisasi LDL dengan Diagnosis	77

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

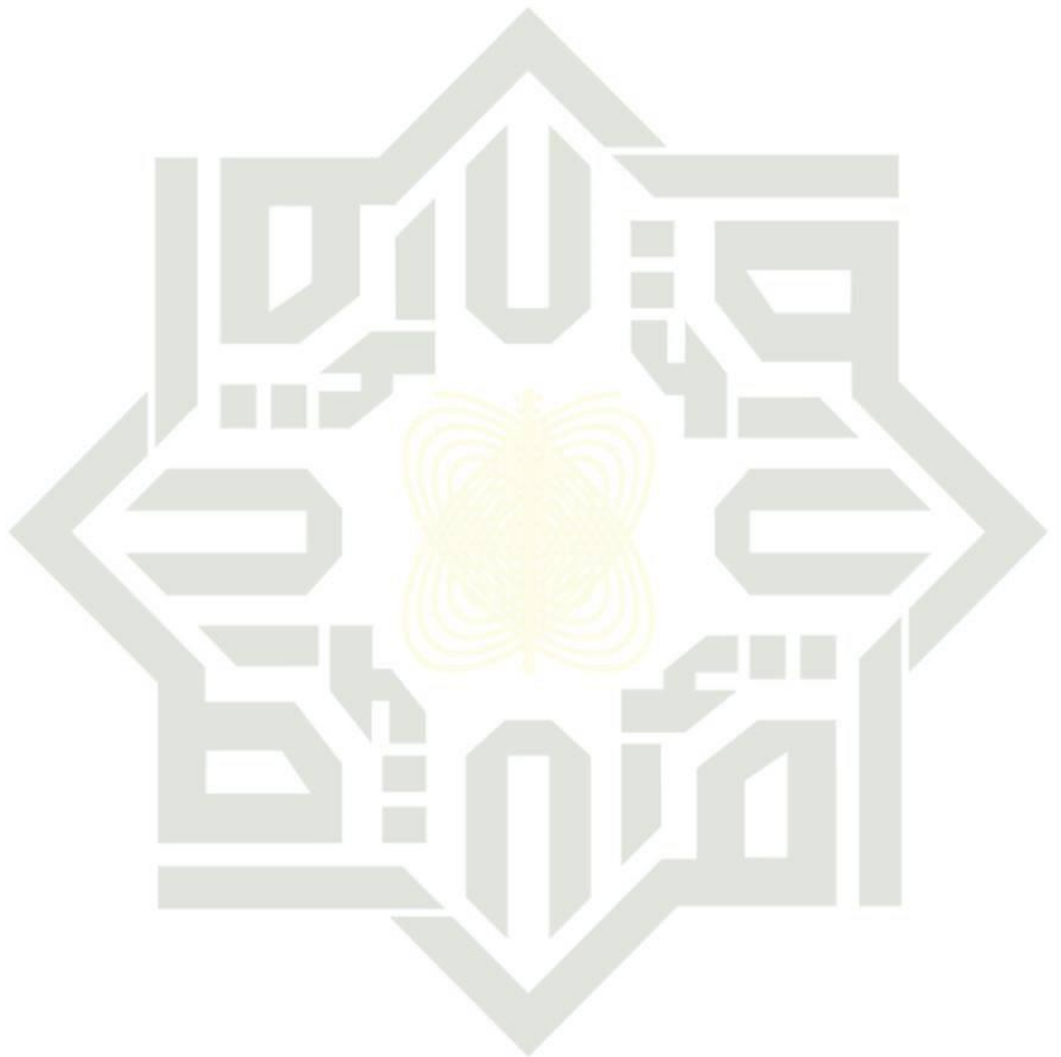
b. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Gambar 32 Visualisasi <i>Systolic</i> dengan Diagnosis	78
Gambar 33 Visualisasi <i>Diastolic</i> dengan Diagnosis.....	78
Gambar 34 Persebaran Data	79



UIN SUSKA RIAU



DAFTAR TABEL

Tabel 1 <i>Confusion Matrix</i>	11
Tabel 2 Klasifikasi Hipertensi	13
Tabel 3 Diagnosi Diabetes Melitus Pada Pemeriksaan Laboratorium	14
Tabel 4 Kadar Kolesterol Darah	15
Tabel 5 Penelitian Terkait	15
Tabel 6 Dataset Awal	23
Tabel 7 Contoh Dataset Penyakit <i>Stroke</i>	30
Tabel 8 Deskripsi Atribut Dataset Awal	31
Tabel 9 Atribut yang digunakan untuk Analisis	33
Tabel 10 Contoh Transformasi data	34
Tabel 11 Dataset Sebelum Normalisasi	35
Tabel 12 Dataset Setelah Normalisasi	37
Tabel 13 Tabel Dataset Testing	38
Tabel 14 Tabel Dataset Training	38
Tabel 15 Hasil Pengujian	62
Tabel 16 Hasil Pengujian Semua Fitur	65
Tabel 17 Hasil Pengujian <i>Feature Selection</i>	68
Tabel 18 Hasil Analisa Pengujian	80

© Hak cipta dimiliki UIN Suska Riau State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



DAFTAR RUMUS

© Hak cipta	10
© Hak cipta	10
© Hak cipta	10
© Hak cipta	11
© Hak cipta	11
© Hak cipta	12
© Hak cipta	12
© Hak cipta	12
© Hak cipta	13

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Stroke adalah penyakit ketiga paling umum di dunia yang menyebabkan kecacatan dan menyebabkan gangguan fungsi saraf karena gumpalan darah yang mengalami penyumbatan (Rahayu & Yamasari, 2024). Menurut *World Health Organization* (WHO), *stroke* adalah penyakit yang berhubungan dengan fungsi saraf. Beberapa bagian tubuh, seperti wajah, kaki, atau tangan, mungkin tampak kesemutan, yang mungkin sering hilang atau tidak, dan mati rasa disatu sisi tubuh (Ismafillah et al., 2023).

Faktor yang memiliki pengaruh yang lebih besar dari pada banyak penyebab lainnya yaitu tekanan darah tinggi (Rahayu & Yamasari, 2024). Menurut data yang dikumpulkan oleh Badan Riset Kesehatan Indonesia, Sekitar 21,1% kematian di Indonesia disebabkan oleh *stroke* (Ismafillah et al., 2023).

Pengembangan model prediktif yang efektif untuk menganalisis data kesehatan dalam era medis yang terus berkembang dengan cepat telah menjadi perhatian besar. Algoritma *K-nearest neighbors* (KNN), yang sering digunakan untuk regresi dan klasifikasi data medis, adalah teknik yang banyak digunakan dalam dunia medis (Martono & Sulistianingsih, 2024). Metode KNN mengklasifikasikan data berdasarkan data latih (training data). Ini adalah bagian dari pembelajaran terawasi (supervised learning), di mana data terbaru dikategorikan berdasarkan kedekatan jumlah kategori yang ada. Dengan kata lain, algoritma KNN mengklasifikasikan data baru berdasarkan jarak datum dan kedekatannya dengan data lain (Setiawan, 2022).

Menurut penelitian yang dilakukan oleh (Zuriati & Qomariyah, 2022) bahwa Algoritma KNN telah terbukti mampu memprediksi sampel uji baru berdasarkan jarak geometri. Jumlah 5110 data yang digunakan termasuk gender, usia,

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

hypertension, penyakit jantung, status menikah, berat badan, jenis pekerjaan, jenis tempat tinggal, tingkat glukosa rata-rata, status merokok, dan kelompok stroke. Proses penelitian terdiri dari pengumpulan data, proses preprocessing, pembagian data, penerapan algoritma KNN, dan evaluasi kinerja KNN melalui penggunaan matrix confusion dan penghitungan akurasi. Algoritma KNN terbaik memiliki nilai $K=5$ dan akurasi 93.54%.

Peneliti sebelumnya telah ada penelitian tentang perbandingan jarak pada Algoritma KNN oleh (Martono dan Sulistianingsih, 2024) dimana Studi ini menggunakan berbagai metode penghitungan jarak, termasuk jarak euclidean, jarak Manhattan, jarak Minkowski, cosine, dan koefisien korelasi Pearson. Data yang digunakan berasal dari pasien diabetes. Metode evaluasi yang digunakan termasuk presisi, sensitivitas, skor F1, dan akurasi. Hasil penelitian menunjukkan bahwa pengukuran jarak dengan Manhattan Distance lebih baik daripada metode pengukuran lainnya; memiliki nilai presisi 0,952, nilai sensitivitas 0,950, nilai skor F1 0,939, dan nilai akurasi 0,950. Sebaliknya, pengukuran jarak dengan jarak euclidean dan Minkowski secara konsisten memberikan hasil yang sama untuk presisi, sensitivitas, F1-Score, dan akurasi, masing-masing 0,932. Berbeda dengan metode Cosine Similarity, yang memiliki akurasi yang tinggi sebesar 0,905, menunjukkan nilai yang lebih rendah untuk ukuran evaluasi seperti presisi, sensitivitas, dan skor F1, masing-masing 0,865, 0,905, dan 0,881. Sedangkan pengukuran jarak dengan Pearson Correlation Coefficient menunjukkan kinerja yang lebih rendah untuk presisi, sensitivitas, dan skor F1, masing-masing 0,544, 0,517, dan 0,881.

Algoritma K-nearest neighbor adalah algoritma yang digunakan untuk mengklasifikasikan objek baru yang telah dipilih berdasarkan kumpulan atribut dan sampel pelatihan, dengan hasil data sampel uji yang baru diklasifikasikan berdasarkan mayoritas kategori pada K-nearest neighbor (Mujammil et al., 2021). Ini dari klasifikasi ini terutama bergantung pada pengukuran jarak atau kesamaan antara contoh yang diuji dan contoh pelatihan. Ini berarti bahwa pengklasifikasi KNN yang menggunakan salah satu dari 10 jarak teratas ada hasil menunjukkan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

bahwa beberapa jarak kurang terpengaruh dibandingkan dengan jarak lainnya (Prasath et al., n.d.).

Berdasarkan latar belakang diatas maka penulis melakukan penelitian tugas akhir yang berjudul Perbandingan Metode Pengukuran Jarak pada Kinerja Algoritma *K-nearest Neighbor* dalam Klasifikasi Penyakit *Stroke*. Data penelitian ini digunakan sebanyak 15.000 data. Perhitungan jarak yang digunakan pada penelitian ini yaitu perhitungan Jarak *Manhattan*, *Euclidean*, *Minkowski*, *Chebyshev* dan *Cosine*.

1.2 Rumusan Masalah

Berdasarkan konteks permasalahan yang telah dijelaskan, maka rumusan masalah penelitian ini adalah:

- 1 Bagaimana perbandingan perhitungan jarak (*Euclidean*, *Manhattan*, *Minkowski*, *Chebyshev* dan *Cosine*) pada metode klasifikasi *K-nearest neighbor* (KNN) pada klasifikasi penyakit *stroke*?
- 2 Bagaimana hasil akurasi tertinggi dari perbandingan evaluasi beberapa pengujian menggunakan metode *K-nearest neighbor* (KNN) pada klasifikasi penyakit *stroke*?

1.3 Batasan Masalah

Untuk menetapkan luasnya pembahasan suatu masalah maka diperlukan pembatasan masalah. Penulis membatasi cakupan topik yang akan dibahas adalah:

1. Analisis data saat ini menggunakan *Kaggle Dataset* yang diambil pada pada Bulan Mei 2024 yang mana atribut fitur telah difilter oleh dr. Andre Lucas, Sp.S, Dokter Spesialis Saraf di RS Awal Bros Pekanbaru agar dapat mengetahui atribut apa saja yang dapat mempengaruhi *stroke*, sehingga diperoleh data dan atribut sebanyak 15000 data, 14 atribut yang terdiri dari 13 atribut fitur dan 1 label kelas.
2. Atribut dari situs *Kaggle Dataset* yang telah dipilih oleh dokter spesialis saraf RS Awal Bros Pekanbaru diperoleh sebanyak 14 atribut yaitu *Age*, *Gender*, *Hypertension*, *Heart Disease*, *Average Glucose Level*, *Body*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Mass Index (BMI), Smoking Status, Alcohol Intake, Dietary Habits, HDL, LDL, Systolic, Diastolic, Diagnosis.

3. Penyakit *stroke* yang akan diteliti menggunakan 2 kelas, terdiri dari menderita *Stroke* dan Tidak *Stroke*.
4. Perhitungan jarak yang digunakan yaitu metode perhitungan jarak (*Euclidean, Manhattan, Minkowski, Chebyshev dan Cosine*).

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Membandingkan metode perhitungan jarak (*Euclidean, Manhattan, Minkowski, Chebyshev dan Cosine*) pada metode klasifikasi *K-nearest neighbor* (KNN) untuk mengklasifikasi terhadap penyakit *stroke*.
2. Memperoleh hasil akurasi tertinggi dari perbandingan evaluasi beberapa pengujian menggunakan metode *K-nearest neighbor* (KNN) pada klasifikasi penyakit *stroke*?

1.5 Manfaat Penelitian

Manfaat penelitian penyakit *stroke* menggunakan metode *K-nearest neighbor* (KNN) dalam klasifikasi *stroke* antara lain :

1. Meningkatkan akurasi klasifikasi penyakit *stroke*. Berdasarkan perbandingan perhitungan jarak *Euclidean, Manhattan, Minkowski, Chebyshev dan Cosine* menggunakan metode *K-nearest neighbor* (KNN).
2. Penelitian ini kedepannya menjadi sumber referensi bagi penelitian berikutnya, khususnya di bidang Jaringan Saraf Tiruan dan penerapan algoritma *K-nearest neighbor* (KNN) untuk meningkatkan akurasi dengan teknik yang berbeda ataupun sama.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 2

KAJIAN PUSTAKA

2.1 Data Mining

Data mining adalah proses mengekstraksi dan mengidentifikasi informasi penting dan terkait dari berbagai database besar dengan menggunakan matematika, statistik, kecerdasan buatan, dan pembelajaran mesin (Widaningsih, 2019). Data mining, juga disebut *Knowledge Discovery in Databases* (KDD), mencakup semua proses mengubah sekumpulan data biasa menjadi pengetahuan yang bermanfaat, dan bertujuan untuk menyelesaikan suatu masalah dengan menggunakan model matematika untuk menemukan pola dari data yang sudah ada sebelumnya (Rahel Lina Simanjuntak et al., 2023).

Penelitian yang dilakukan oleh (Azhar et al., 2022) menunjukkan bahwa Data mining dapat digunakan untuk memprediksi pasien stroke; dalam penelitian, dataset dibagi menjadi dua bagian: data pelatihan dan data pengujian dengan menggunakan validasi split. Menurut hasil pengujian yang dilakukan dalam penelitian ini, algoritma dengan tingkat akurasi tertinggi adalah *Logistic Regression*, *Random Forest*, SVM, dan KNN. Nilai akurasi mereka adalah 98,63%.

2.1.1 Knowledge Discovery in Database

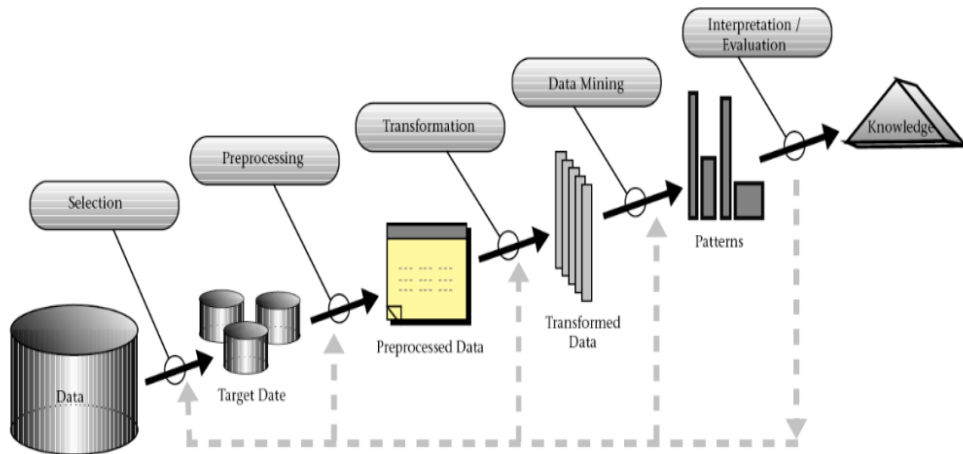
Pada penelitian (Juppren dan Arnomo, 2021) membahas tentang *Knowledge Discovery In Databases* (KDD) adalah sistem untuk menemukan informasi dalam database yang tersedia. Tabel-tabel yang berkorelasi ditemukan dalam database. Sebuah basis pengetahuan dibuat dari data yang ditemukan dalam database untuk digunakan sebagai pertimbangan saat membuat keputusan.

Knowledge Discovery in Database (KDD) dan data mining dikolaborasi oleh pihak-pihak berkepentingan sesuai kebutuhan pada kondisi tertentu untuk menjabarkan operasi penambahan informasi yang mungkin tetapi belum terjadi pada sebuah basis data yang cukup besar. Meskipun konsep KDD dan data mining tidak sama, keduanya berkolaborasi satu sama lain, dan data mining adalah salah

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sebagai satu hierarki prosedur KDD. Berikut prosedur KDD yang dijabarkan seperti di bawah ini:



Gambar 1 Tahapan KDD

Berikut ini adalah penjelasan mengenai proses tahapan KDD (Juppren & Arnomo, 2021):

1. Data Selection (Seleksi Data)

Untuk memulai melakukan seleksi pada tahap Penemuan Pengetahuan dalam Database. Setelah melakukan seleksi, hasilnya disimpan secara terpisah dari basis data operasional, yang akan digunakan sebagai sumber data mining.

2. Pre-processing / Cleaning

Setelah itu, tahap pembersihan harus diperhatikan. Proses pembersihan mencakup menghapus data yang tidak relevan, memeriksa data yang tidak konsisten, dan memperbaiki data yang dianggap salah atau tidak sesuai, seperti kesalahan pencetakan. Selanjutnya, data "diperkaya" dengan data baru yang dikombinasikan dengan data yang telah tersedia sebelumnya untuk digunakan dalam *Knowledge Discovery in Database* (KDD), seperti informasi eksternal. Proses ini dikenal sebagai enrichment.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Data Transformasi

Pencarian corak tertentu dalam data yang telah dipilih sebelumnya melalui kaidah atau sistem yang ada dikenal sebagai data mining. Ada banyak metode dan algoritma yang dapat digunakan dalam data mining. Metode dan algoritma ini dipilih berdasarkan tujuan dan kebutuhan pengolahan data untuk melakukan *Knowledge Discovery in Database* (KDD).

4. Data Mining

Selanjutnya disebut juga sebagai proses penambangan data, yang berarti proses utama yang menggunakan metode untuk menemukan pengetahuan berharga yang tersembunyi dari data.

5. Interpretation / Evaluation

Corak yang terbentuk selama tahap transformasi dan pengolahan data harus menghasilkan sebuah hasil yang memudahkan semua pihak yang memiliki kepentingan untuk membuat keputusan. Karena itu, ini adalah langkah terakhir, interpretasi. Pada tahap ini, informasi dan pola dievaluasi untuk mengetahui apakah data bertentangan dengan kenyataan yang sebenarnya atau apakah jawaban sementara yang telah dirumuskan.

2.2 Klasifikasi

Salah satu pendekatan data mining adalah klasifikasi. Klasifikasi sendiri adalah suatu proses menempatkan item atau ide tertentu ke dalam rangkaian kategori yang ditampilkan berdasarkan item yang digunakan. Jaringan saraf tiruan, naive bayes, support vector machine, pohon keputusan, dan fuzzy adalah beberapa teknik klasifikasi, pohon keputusan adalah yang paling banyak digunakan. Klasifikasi sendiri terdiri dari dua tahap: klasifikasi dan pembelajaran. Pada tahap pembelajaran, algoritma klasifikasi menggunakan data pelatihan untuk membuat model klasifikasi. Selain itu, fase pembelajaran dapat digambarkan sebagai fase membangun fungsi atau pemetaan, di mana y adalah kelas yang diprediksi dan x adalah tupel yang diinginkan kelas tersebut (Mujammil et al., 2021).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.3 K-nearest neighbor (KNN)

Algoritma K-nearest neighbor digunakan untuk mengklasifikasikan data berdasarkan data pembelajaran yang memiliki jarak ketetanggaan yang paling dekat dengan objek penelitian. Tujuan algoritma ini adalah untuk mengklasifikasikan objek penelitian baru berdasarkan nilai masing-masing atribut. Algoritma K-nearest neighbor adalah algoritma yang digunakan untuk mengklasifikasikan objek baru yang telah dipilih berdasarkan kumpulan atribut dan sampel pelatihan, dengan hasil data sampel uji yang baru diklasifikasikan berdasarkan mayoritas kategori pada K-nearest neighbor (Mujammil et al., 2021).

Pada penelitian ini mengklasifikasikan penyakit *stroke* menggunakan metode KNN data terlebih dahulu menentukan nilai K, kemudian data dinormalisasikan menggunakan Normalisasi *Min-max*.

Berikutnya rumus menghitung Normalisasi *Min-max*:

$$X_{norm} = \left(\frac{X - X_{min}}{X_{max} - X_{min}} \right) \dots\dots\dots(1)$$

Keterangan :

X_{norm} = nilai yang sudah di normalisasi

X = nilai asli dari suatu atribut

X_{max} = nilai max dari atribut

X_{min} = nilai min dari atribut

Kemudian menghitung pengukuran jarak *Euclidean*, *Manhattan*, *Minkowski*, *Chebyshev* dan *Cosine*. Setelah mendapatkan perhitungan jarak dilanjutkan dengan mengurutkan hasil jarak terkecil ke terbesar kemudian mencari mayoritas kelas dari nilai K tetangga dan dijadikan hasil prediksi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Berikutnya rumus menghitung jarak *Euclidean*, *Manhattan*, *Minkowski*, *Chebyshev* dan *Cosine* :

Rumus Jarak *Euclidean*

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots\dots\dots(2)$$

Keterangan:

- X₁ = data latih
- X₂ = data uji
- i = atribut data d = jarak
- p = dimensi data

Rumus Jarak *Manhattan*

$$d_i = \sum_{i=1}^p |x_{2i} - x_{1i}| \dots\dots\dots(3)$$

Keterangan:

- X₁ = data latih
- X₂ = data uji
- i = atribut data d = jarak
- p = dimensi data

Rumus Jarak *Minkowski*

$$d_{(p,q)} = (\sum_{i=1}^n |p_i - q_i|^p)^{1/p} \dots\dots\dots(4)$$

Keterangan:

d(p,q) = Jarak antara titik p dan q.

p dan q = Dua titik yang ingin dihitung jaraknya. Masing-masing punya beberapa komponen atau fitur, misalnya p=(p₁,p₂,...,p_n) dan q=(q₁,q₂,...,q_n).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

n = Jumlah komponen atau dimensi (fitur) dari tiap titik.

P = Parameter untuk menentukan jenis jarak:

Jika $p=1$ = Menjadi jarak Manhattan atau "jarak taksi".

Jika $p=2$ = Menjadi jarak Euclidean atau "jarak lurus".

Jika $p \rightarrow \infty$ = Menjadi jarak Chebyshev, yang mengambil nilai maksimum perbedaan komponen.

Rumus Jarak Chebyshev

$$d_{(p,q)} = (\max|p_i - q_i|) \dots\dots\dots(5)$$

Keterangan:

$d(p,q)$ = Jarak *Chebyshev* antara titik p dan q .

$(p$ dan $q)$ = Dua titik dalam ruang n -dimensi yang ingin dihitung jaraknya.

$|p_i - q_i|$ = Nilai absolut dari perbedaan antara komponen dari p dan q .

Max = Mengambil nilai terbesar dari perbedaan semua koordinat.

Rumus Jarak Cosine

$$d_{cosine}(x, y) = 1 - \frac{\sum_{i=1}^n x \cdot y}{\sqrt{\sum_{i=1}^n x^2} \cdot \sqrt{\sum_{i=1}^n y^2}} \dots\dots\dots(6)$$

Keterangan:

x = Data Uji

y = Data Latih

2.4 Confusion Matrix

Confusion Matrix merupakan salah satu cara untuk menganalisis kinerja model klasifikasi (Nikmatun & Waspada, 2019). *Confusion Matrix* dapat dilihat pada sebagai berikut.

Tabel 1 *Confusion Matrix*

<i>Confusion Matrix</i>		Nilai Prediksi	
		Positif	Negative
Nilai sebenarnya	Positif	TP	FN
	Negative	FP	TN

Untuk mengetahui tingkat akurasi, kita harus membandingkan jumlah data yang benar diklasifikasikan dengan jumlah total data. Cara perhitungannya dapat dilihat dibawah ini.

$$Akurasi = \frac{TP+TN}{P+N} \dots\dots\dots(7)$$

Precision adalah ukuran ketepatan jika prediksi data positif. Nilai precision dapat dilihat dibawah ini.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(8)$$

Namun, recall menunjukkan kelengkapan. Jumlah data yang diprediksi positif sama dengan jumlah data sebenarnya yang bernilai positif. Nilai recall dapat dilihat dibawah ini.

$$Recal = \frac{TP}{TP+FN} \dots\dots\dots(9)$$

Keterangan:

- = nilai true positives
- = nilai true negatives
- = nilai false positives
- = nilai false negatives
- = jumlah data

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.5 Stroke

Stroke adalah penyakit otak yang disebabkan oleh tekanan darah tinggi. Jika terjadi perdarahan karena tekanan intrakranial yang tinggi dan keluarnya embolus dari pembuluh darah non serebral, *stroke* dapat terjadi kapan saja. Tekanan darah tinggi dapat menyebabkan hipertensi. Perubahan tekanan darah jangka panjang juga dapat menyebabkan *stroke*, yang menghalangi aliran darah, memungkinkan serangan dan kematian sel-sel saraf otak. Karena fakta bahwa penyakit ini muncul secara tiba-tiba dan membutuhkan penanganan segera, itu diklasifikasikan sebagai penyakit serebrovaskular (CVD). Sumber daya yang terbatas dan ahli saraf yang terlatih masih menjadi tantangan bagi setiap negara untuk mengatasi angka *stroke* (Caputri et al., 2022).

Baik di negara maju maupun negara berkembang, *stroke* adalah penyebab kematian ketiga tertinggi di dunia, setelah penyakit jantung koroner dan kanker. Menurut data *World Stroke Organization*, setiap tahun terjadi 13,7 juta kasus baru penyakit *stroke*, dan sekitar 5,5 juta kematian disebabkan oleh *stroke* (Setiawan et al, 2021).

Pada buku tentang “*Awas Stroke*” yang ditulis oleh (Pinzon, n.d.) Buku ini menunjukkan bahwa Faktor risiko *stroke* dapat dibagi menjadi dua kategori: faktor risiko yang dapat diubah dan faktor risiko yang tidak dapat diubah. Faktor risiko yang tidak dapat diubah termasuk usia, jenis kelamin, ras, riwayat keluarga, dan *stroke* sebelumnya. *Stroke* lebih mungkin terjadi pada orang yang lebih tua. *Stroke* dapat terjadi pada setiap usia, tetapi lebih dari 70% kasus terjadi pada orang di atas 65 tahun. Disebabkan oleh fakta bahwa laki-laki memiliki tingkat kejadian faktor risiko *stroke* yang lebih tinggi, seperti hipertensi, laki-laki lebih rentan terhadap *stroke*. Sangat penting untuk memahami faktor risiko yang dapat diubah ini untuk *stroke*. Penanganan berbagai faktor risiko ini bertujuan untuk mengurangi risiko *stroke*. Hipertensi, diabetes, merokok, dan dislipidemia adalah faktor risiko utama *stroke*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.5.1 Faktor Risiko Stroke

Faktor risiko penyakit *stroke* disebutkan oleh dr. Andre Lucas, Sp.S, Dokter Spesialis Saraf di Pekanbaru diantaranya:

1. Usia
Semakin tua usia seseorang akan semakin mudah terkena *stroke*. *Stroke* dapat terjadi pada semua usia, namun > 70% kasus *stroke* terjadi pada usia diatas 65 tahun(Pinzon, n.d.).
2. Jenis Kelamin
Laki-laki lebih mudah terkena *stroke*. Hal ini dikarenakan lebih tingginya angka kejadian faktor risiko *stroke* (misalnya: hipertensi) pada laki-laki(Pinzon, n.d.).
3. Hipertensi
Hipertensi merupakan faktor risiko *stroke* yang paling utama dan konsisten. Kajian Vasan, *et al.*, (2001) memperlihatkan bahwa peningkatan tekanan darah sistolik dan diastolik secara konsisten meningkatkan risiko kejadian *stroke* dan penyakit kardiovaskuler lainnya(Pinzon, n.d.).

Tabel 2 Klasifikasi Hipertensi

Klasifikasi	Tekanan darah sistolik	Tekanan darah diastolic
Normal	<120	<80
Pra hipertensi	120-139	80-89
Hipertensi tingkat 1	140-159	90-99
Hipertensi tingkat 2	>160	>100

4. Penyakit Jantung
Gangguan irama jantung merupakan faktor risiko *stroke* kardioembolik yang utama. Fibrilasi atrium secara signifikan meningkatkan risiko *stroke*(Pinzon, n.d.).
5. Tingkat Glukosa
Diabetes dijumpai pada 15%-20% populasi usia dewasa. Diabetes merupakan salah satu faktor risiko *stroke* iskemik yang utama. Diabetes

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

akan meningkatkan risiko *stroke* 2 kali lipat. Peningkatan kadar gula darah berhubungan lurus dengan risiko *stroke* (semakin tinggi kadar gula darah, semakin mudah terkena *stroke*)(Pinzon, n.d.).

Tabel 3 Diagnosi Diabetes Melitus Pada Pemeriksaan Laboratorium

	Normal	GTT	DM
Gula darah puasa	<110	110-125	≥126
Dua jam setelah beban glukosa	<140	140-200	≥200

GTT : Gangguan toleransi glukosa

DM : Diabetes mellitus

6. **Obesitas**
 Seseorang dengan berat badan berlebih memiliki risiko yang tinggi untuk menderita *stroke*. Penelitian Oki, dkk (2006) menyimpulkan bahwa seseorang dengan indeks massa tubuh ≥ 30 memiliki risiko *stroke* 2,46 kali dibanding yang memiliki indeks massa tubuh < 30 (Pinzon, n.d.).
7. **Merokok**
 Berbagai penelitian menghubungkan kebiasaan merokok dengan peningkatan risiko penyakit pembuluh darah (termasuk *stroke*). Merokok memacu peningkatan kekentalan darah, pengerasan dinding pembuluh darah, dan penimbunan plak di dinding pembuluh darah(Pinzon, n.d.).
8. **Penyalahgunaan obat / alkohol**
 Menghindari *stroke* dengan menghentikan kebiasaan merokok, minum alcohol, mengkonsumsi obat-obatan, serta memperbanyak olahraga(Setiawan et al, 2021).
9. **Kebiasaan Jenis Diet**
 Diet yang tidak sehat mengkonsumsi makanan atau minuman yang manis dapat menyebabkan peningkatan resiko *stroke*(Utama & Nainggolan, 2022).
10. **Tingkat Kolesterol**
 Tubuh menghasilkan kolesterol dari dua jenis utama: kolesterol LDL dan kolesterol HDL. Kolesterol LDL, yang dikenal sebagai "kolesterol jahat",

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

membawa kolesterol dari hati ke dalam sel, menyebabkan penimbunan kolesterol di dalam sel. Hal ini menyebabkan pengerasan dinding pembuluh darah arteri, atau atherosklerosis, yang membahayakan ginjal, jantung, dan otak, dan meningkatkan risiko stroke pada otak (Pinzon, n.d.).

Tabel 4 Kadar Kolesterol Darah

Kadar Kolesterol	Risiko	Rekomendasi
<200 mg/dL	Rendah	<ul style="list-style-type: none"> – Cek berkala – Cek factor risiko lain
200-239 mg/dL	Sedang	<ul style="list-style-type: none"> – Perubahan pola hidup – Cek berkala minimal 1 tahun
>240 mg/dL	Tinggi	<ul style="list-style-type: none"> – Perubahan pola hidup – Hubungi dokter

2.6 Penelitian Terkait

Penelitian ini membahas penelitian yang berkaitan dengan latar belakang masalah dari penelitian ini. Di bawah tabel penelitian terkait kajian, terdapat beberapa kajian yang dijelaskan pada tabel 5.

Tabel 5 Penelitian Terkait

No	Peneliti	Judul	Kesimpulan
1	(Setiawan et al., 2024)	Komparasi Algoritma <i>K-nearest neighbor</i> (KNN), Support Vector Machine (SVM), dan Decision Tree Dalam Klasifikasi Penyakit <i>Stroke</i>	Hasil penelitian menunjukkan bahwa tiga algoritma yang digunakan, yaitu Decision Tree, K-NN, dan SVM, memiliki akurasi terbaik. Algoritma KNN memiliki akurasi 94%, SVM 95%, dan Decision Tree memiliki akurasi 92%.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2	(Martono & Sulistianingsih, 2024)	Perbandingan Matriks jarak pada Algoritma K-NN untuk Prediksi Penyakit Diabetes	Sebuah analisis tentang bagaimana metrik pengukuran jarak bekerja menunjukkan hasil yang menarik. Dibandingkan dengan Manhattan Distance, Euclidean Distance menunjukkan hasil yang sangat baik dengan presisi, sensitivitas, dan skor F1 sekitar 0,95, dan akurasi 0.950. Minkowski Distance menunjukkan hasil yang serupa dengan Euclidean Distance, menegaskan konsistensinya dengan presisi, sensitivitas, dan skor F1 sekitar 0.930 hingga 0.940, dan akurasi 0.950. Untuk <i>cosine</i> meskipun akurasinya sekitar 0.905, <i>Cosine Similarity</i> menunjukkan skor presisi, sensitivitas, dan F1 yang lebih rendah dibandingkan dengan metrik lainnya, dengan skor antara 0.865 dan 0.881. Pearson Correlation Coefficient juga menunjukkan skor yang lebih rendah dalam presisi, sensitivitas, dan F1, terutama dengan skor presisi hanya sekitar 0.544.
3	(Uni Umamaton Nysa et al., 2023)	Perbandingan Jarak Euclidean, Manhattan, Chebyshev Pada Klasifikasi Status Gizi Balita	Data status gizi Posyandu Dusun Pilang dibandingkan dengan uji penentuan status gizi perbandingan jarak Euclidean, Manhattan, dan Chebyshev menggunakan metode KNN. Hasilnya menunjukkan bahwa 25 data serupa dan 7 tidak serupa.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		Menggunakan Metode <i>K-nearest neighbors</i> (KNN)	Hasil-hasil di atas menunjukkan bahwa Model KNN dapat mengklasifikasikan sampel dengan baik sebagai "Normal", tetapi dia mengalami kesulitan membedakan antara kelas "Gemuk" dan "Kurus".
4	(Rahel Lina Simanjuntak et al., 2023)	Komparasi Algoritma KNN dan SVM dalam Memprediksi Penyakit <i>Stroke</i>	Berdasarkan hasil uji coba yang dilakukan menggunakan algoritma Support Vector Machine (SVM) dan K-nearest neighbors (KNN) dalam bahasa pemrograman Python di Google Colab, dapat disimpulkan bahwa SVM memberikan kinerja dan akurasi terbaik untuk memprediksi penyakit stroke. Algoritma SVM mencapai akurasi 100%, sementara KNN mencapai akurasi hanya 97%, perbedaan 3% lebih rendah dari SVM. Oleh karena itu, algoritma Support Vector Machine (SVM)
5	(Akmal et al., 2023)	Perbandingan Metode Algoritma Naïve Bayes dan <i>K-nearest neighbors</i> Untuk Klasifikasi Penyakit <i>Stroke</i>	Hasil perbandingan terhadap diagnosa penyakit stroke menggunakan klasifikasi metode Naïve Bayes dan K-nearest neighbors menunjukkan bahwa algoritma K-nearest neighbors memiliki nilai ketepatan Performance Vector terbaik sebesar 94.36%, sedangkan algoritma Naïve Bayes memiliki nilai ketepatan terendah sebesar 90.10%.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6	(Zuriati & Qomariyah, 2022)	Klasifikasi Penyakit <i>Stroke</i> Menggunakan Algoritma <i>K-nearest neighbor</i> (KNN)	Berdasarkan hasil penelitian, algoritma KNN telah digunakan untuk mengklasifikasikan kumpulan data penyakit stroke dengan sukses. Nilai akurasi tertinggi dicapai pada nilai $k = 5$, dengan komposisi perbandingan data pelatihan dan pengujian 90%:10% dan nilai akurasi 93.54%.
7	(Azhar et al., 2022)	Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit <i>Stroke</i>	Hasil penelitian menunjukkan bahwa Logistic Regression memiliki tingkat akurasi tertinggi sebesar 75.65% pada data yang diseimbangkan, sementara Random Forest, SVM, dan KNN memiliki tingkat akurasi tertinggi sebesar 98.63% pada data yang belum diseimbangkan. Untuk nilai akurasi algoritma data mining ada dua data. Data yang diimbangi memiliki nilai akurasi Random Forest 72,17%, SVM 76,52%, dan KNN 72,61%, dan data yang tidak diimbangi memiliki nilai akurasi Random Forest 72,17%, SVM 76,52%, dan KNN 98,63%. Berdasarkan pengecekan akurasi dengan confusion matrix di atas, kita dapat mengetahui perbandingan jumlah TRUE POSITIF, TRUE NEGATIF, FALSE NEGATIF, dan FALSE NEGATIF dari kedua model. Studi kasus ini menunjukkan bahwa model yang memprediksi lebih

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

			banyak pasien yang mengalami stroke (TRUE POSITIF) lebih baik karena model dapat memprediksi kecenderungan pasien yang memiliki kemungkinan besar mengalami stroke, meskipun faktanya mereka belum didiagnosa stroke.
8	(Sulaeman, 2022)	Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit <i>Stroke</i> Support Vector Machine Algorithm Analysis In <i>Stroke</i> Disease Classification	Dataset Kaggle, yang dirilis pada tahun 2021, digunakan untuk melakukan penelitian dan pengujian. Untuk mengklasifikasikan data, Support Vector Machine (SVM) digunakan. Penelitian ini menggunakan dua cara untuk menguji data: metode pengujian data tidak seimbang untuk data tidak seimbang dan metode klasifikasi umum untuk data seimbang. Dalam SVM, dua jenis kernel digunakan: kernel linear dan polinomial. Hasil pengujian menunjukkan bahwa kernel linear memiliki akurasi tertinggi sebesar 76% pada data tidak seimbang dan kernel polynomial sebesar 80% pada data seimbang. Untuk data seimbang, kernel linear memiliki akurasi tertinggi sebesar 77% dan kernel polynomial sebesar 76%.
9	(Aziz et al., 2021)	Analisa Perbandingan	Berdasarkan analisis uji coba yang melibatkan beberapa skenario, dapat

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		Algoritma Euclidean Dan Manhattan <i>Distance</i> Dalam Identifikasi Wajah	disimpulkan bahwa algoritma penghitungan jarak Euclidean memiliki presentase tertinggi sebesar 100%, dibandingkan dengan algoritma penghitungan jarak Manhattan yang hanya memiliki presentase sebesar 85%. Semakin banyak data latihan yang digunakan, semakin akurat hasil uji coba untuk algoritma Euclidean Distance. Maksimal akurasi diperoleh saat setiap karyawan memiliki minimal delapan gambar data latih.
10	(Wijaya et al., 2020)	Perbandingan Akurasi <i>Euclidean Distance</i> , <i>Minkowski Distance</i> , dan <i>Manhattan Distance</i> pada Algoritma <i>KMeans Clustering</i> berbasis <i>Chi-Square</i>	Untuk pelabelan kluster status disparitas kebutuhan Guu, perbandingan akurasi metode pengukuran jarak (<i>Euclidean</i> , <i>Manhattan</i> , dan <i>Minkowski</i>) telah dilakukan. Hasilnya menunjukkan tingkat akurasi yang tinggi sebesar 84.47% untuk metode <i>Euclidean</i> , 83.85% untuk metode <i>Manhattan</i> , dan 83.85% untuk metode <i>Minkowski</i> . Oleh karena itu, dapat disimpulkan bahwa metode <i>Euclidean</i> adalah yang terbaik untuk diterapkan dalam algoritma clustering <i>KMeans</i> .
1	(Wahyono et al., 2020)	Perbandingan penghitungan jarak pada k-	Pada sebagian besar ukuran K, jarak <i>Euclidean</i> dan <i>Minkowski</i> dengan p sebesar 1,5 memiliki akurasi terbaik.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		nearest neighbour dalam klasifikasi data tekstual	Sebaliknya, jarak Chebyshev adalah perhitungan jarak terburuk berdasarkan akurasi. Karena jumlah K yang meningkat menghasilkan lebih banyak data yang tidak memiliki kelas yang sama daripada data dengan kelas yang tepat, menambah jumlah K pada setiap perhitungan jarak menurunkan akurasi classifier. Jumlah K paling efektif adalah $K = 3$.
12	(Nishom, 2019)	Aplikasi Klasifikasi Konsentrasi Studi Mahasiswa Berdasarkan Nilai dan Minat Menggunakan Metode Learning Vector Quantization 2	Akurasi mencapai sebesar 92% berdasarkan hasil klasifikasi dalam menentukan konsentrasi studi siswa berdasarkan nilai dan minatnya.

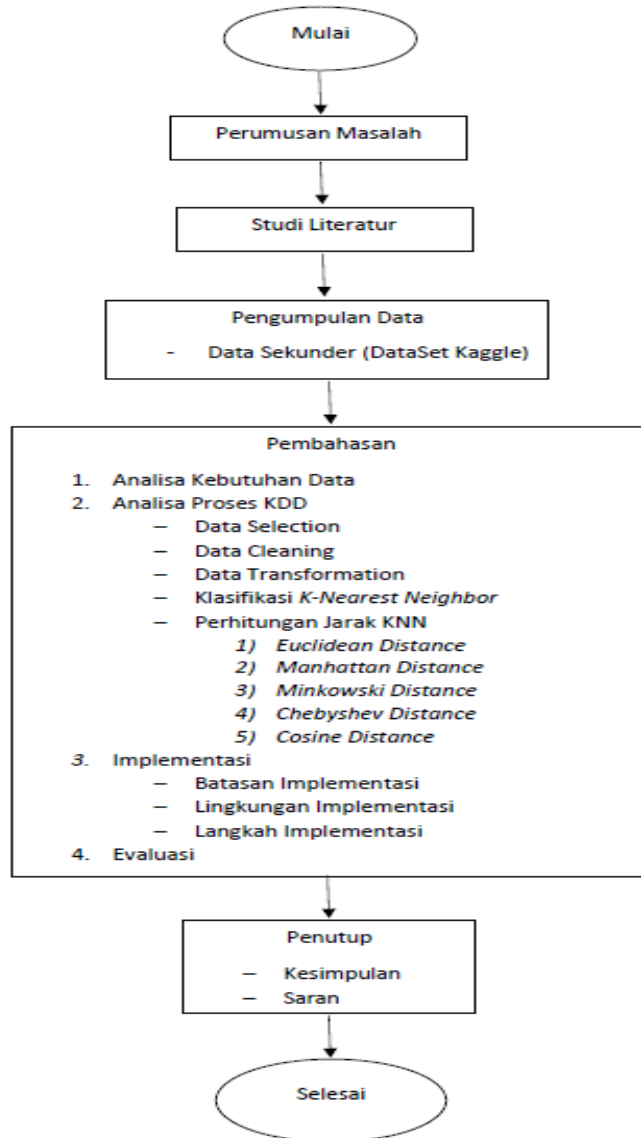
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 3 METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Metodologi penelitian terdiri dari beberapa langkah yang saling terkait. Tujuan penelitian dijelaskan dan diuraikan secara sistematis, jelas, dan teratur dalam metode penelitian. Adapun metodologi penelitian yang dilakukan dapat dilihat pada **Gambar 2**.



Gambar 2 Tahapan Metodologi Penelitian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.2 Perumusan Masalah

Perumusan masalah adalah tahap utama dari metodologi penelitian. Pada penelitian yang akan dilakukan untuk mengidentifikasi masalah bagaimana menerapkan metode *K-nearest neighbor* (KNN) untuk mengklasifikasi terhadap penyakit *stroke* yang terkena *stroke* atau tidak *stroke* dengan membandingkan perhitungan jarak (*Euclidean, Manhattan, Minkowski, Chebyshev dan Cosine*).

3.3 Studi Literatur

Pada tahapan ini penulis melakukan eksplorasi dan mengumpulkan data dari kasus dan referensi terkait baik dari segi topik permasalahan, data dan metode. Referensi penelitian ini dapat melalui beberapa sumber dengan membaca jurnal, artikel dan e-book yang berkaitan dengan penelitian.

3.4 Pengumpulan Data

Dalam pengumpulan data penulis memperoleh sumber data yang akan digunakan dalam penelitian ini melalui metode pengumpulan data skunder. Data penelitian yang di ambil dari platform kaggle. Data dapat diakses melalui [Stroke Prediction \(kaggle.com\)](https://www.kaggle.com/datasets/ahmedmohamed2020/stroke-prediction).

Dataset *stroke* yang akan digunakan dapat di lihat pada tabel menunjukkan dataset awal pada penyakit *stroke*, dataset penyakit *stroke* ini berjumlah 15000 data dengan mempunyai atribut data sebanyak 22 atribut yaitu *Patient ID, Patient Name, Age, Gender, Hypertension, Heart Disease, Marital Status, Work Type, Residence Type, Average Glucose Level, Body Mass Index (BMI), Smoking Status, Alcohol Intake, Physical Activity, Stroke History, Family History od Stroke, Dietary Habits, Blood Pressure Levels, Cholesterol Levels, Symptomps, Diagnosis*).

Tabel 6 Dataset Awal

No	Patient ID	Patient Name	Age	Gender	Hyper tension	Heart Disease	Marital Status	Work Type	Residence Type	...	Diagnosis
1	18153	Mamooty Khurana	56	Male	0	1	Married	Self-employed	Rural	...	Stroke
2	62749	Kaira Subramaniam	80	Male	0	0	Single	Self-employed	Urban	...	Stroke
3	32145	Dhanush Balan	26	Male	1	1	Married	Never Worked	Rural	...	Stroke

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4	6154	Ivana Baral	73	Male	0	0	Married	Never Worked	Urban	...	No Stroke
5	48973	Darshit Jayaraman	51	Male	1	1	Divorced	Self-employed	Urban	...	Stroke
6	29307	Advika Kota	62	Female	0	0	Single	Private	Urban	...	Stroke
7	25525	Elakshi Karan	40	Female	1	0	Married	Private	Urban	...	No Stroke
...
15000	7372	Raghav Handa	72	Female	1	1	Married	Self-employed	Rural	...	Stroke

3.5 Analisa Data

Analisa merupakan tahapan menguraikan detail yang mengandung gambaran penelitian yang akan dilakukan. Berikut adalah tahapan analisa yaitu :

Data Selection

Pada tahap ini dilakukan tahapan seleksi data. Dataset menggunakan Kaggle Dataset yang diambil pada pada Bulan Mei 2024 yang diperoleh data dan atribut sebanyak 15000 data, 22 atribut yang terdiri dari 21 atribut fitur dan 1 label kelas. Kemudian atribut di seleksi oleh dr. Andre Lucas, Sp.S, Dokter Spesialis Saraf di RS Awal Bros Pekanbaru agar dapat mengetahui atribut apa saja yang dapat mempengaruhi *stroke*, atribut yang telah diseleksi diperoleh sebanyak 13 atribut fitur yaitu *Age, Gender, Hypertension, Heart Disease, Average Glucose Level, Body Mass Index (BMI), Smoking Status, Alcohol Intake, Dietary Habits, HDL, LDL, Systolic dan Diastolic*. Dan atribut kelas yaitu *Diagnosis*.

Pre-processing /Cleaning

Pada tahap *cleaning* penulis akan mengatasi *missing value*, data duplikat, data yang tidak konsisten, dan outlier yang ditemukan. Berikut adalah beberapa cara untuk mengatasi masing-masing masalah tersebut:

1. Mengatasi *missing value*

Missing value merupakan data yang hilang atau tidak tersedia dalam dataset. Untuk mengatasi masalah ini, terdapat beberapa metode yang dapat diterapkan, antara lain: menghapus *missing value* atau mengisi *missing value* tersebut dengan nilai tertentu, seperti rata-rata (*mean*), median, atau modus.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Mengatasi data duplikat

Data duplikat adalah baris yang sama yang muncul lebih dari sekali dalam dataset. Untuk mengatasi masalah ini, data duplikat dapat dihapus dari dataset.

3. Mengatasi data yang tidak konsisten

Data yang tidak konsisten dapat berupa format yang berbeda untuk nilai yang sama, kesalahan penulisan, atau data yang berada di luar rentang yang seharusnya. Untuk mengatasi masalah ini, dapat diterapkan beberapa metode, yaitu normalisasi format data, koreksi kesalahan penulisan, dan validasi data.

4. Mengatasi outlier

Outlier adalah nilai yang secara signifikan berbeda dari nilai lainnya dalam dataset. Metode untuk mengatasi outlier antara lain: menghapus outlier, melakukan transformasi data, dan menerapkan winsorizing.

3. *Transformation*

Metode *K-nearest neighbor* (K-NN) merupakan metode yang mampu dipergunakan bila data berupa angka. Proses tranformasi data mengubah atribut data selain angka ke dalam nilai angka, sehingga algoritma ini dapat digunakan untuk mengolah data tersebut. Contoh transformasi data yang digunakan yaitu terdapat pada atribut jenis kelamin, dimana jenis kelamin Laki-laki ditransformasikan menjadi angka 1 dan jenis kelamin Perempuan ditransformasikan menjadi angka 0.

3.5.1 Normalisasi Data

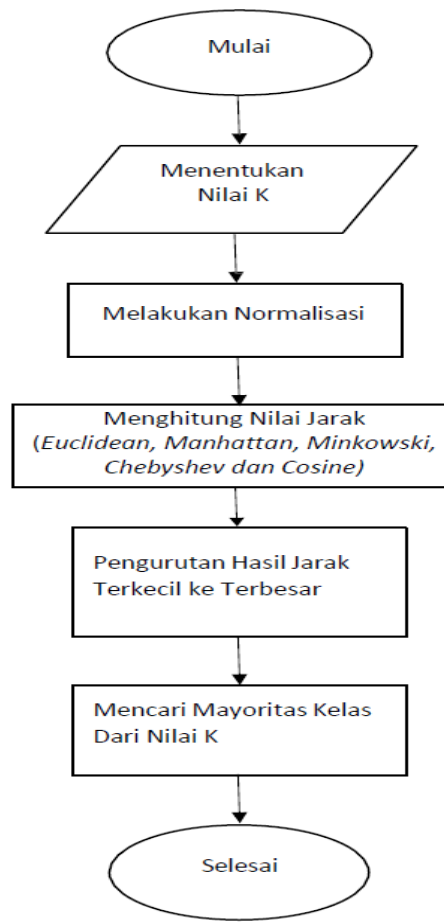
Normalisasi data bertujuan untuk mengatasi nilai dataset agar memiliki skala atau distribusi yang konsisten. Tahapan ini untuk mengukur data dan rentang tertentu, yang mana menggunakan metode *min-max* untuk menentukan pada tahapan ini. Rumus normalisasi dirujuk pada Rumus (1).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.6 *K-nearest neighbor*

Pada tahap data mining, penulis akan menerapkan algoritma *K-nearest neighbor*. Langkah- langkah dalam menerapkan algoritma *K-nearest neighbor* dalam bentuk *flowchart* yaitu:



Gambar 3 *Flowchart* Proses Pelatihan

Dibawah ini adalah prosedur penghitungan menggunakan algoritma *K-nearest neighbor* (KNN), seperti terlihat pada Gambar 3.

- a) Pada tahap pertama ditentukan nilai K. Pemilihan nilai K (jumlah tetangga terdekat yang dipertimbangkan dalam klasifikasi) dapat berupa bilangan ganjil. Ketika K adalah bilangan ganjil, lebih mudah untuk mencapai keputusan mayoritas yang jelas antara dua atau lebih kelas. Misalnya, jika K=3, hasil klasifikasi bisa diputuskan dengan 2 suara terhadap 1 suara.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Bilangan ganjil memberikan kestabilan dalam pengambilan keputusan klasifikasi terutama dalam kasus dua kelas, di mana jumlah suara mayoritas lebih jelas.

- b) Melakukan normalisasi, perhitungan normalisasi menggunakan *Min-max*. Tujuan utama menormalisasi data yakni agar data yang memiliki dimensi kecil tetap dapat mewakili data asli tanpa kehilangan karakteristik data.
- c) Menghitung jarak (*Euclidean, Manhattan, Minkowski, Chebyshev dan Cosine*) data latih dan data uji.
- d) Mengurutkan hasil jarak data terkecil ke terbesar.
- e) Mencari mayoritas kelas dari nilai K tetangga dan dijadikan hasil prediksi. Untuk menghitung *Euclidean* gunakan Rumus (2), *Manhattan* gunakan Rumus (3), *Minkowski* gunakan Rumus (4), *Chebyshev* gunakan Rumus (5), sedangkan untuk menghitung *Cosine* gunakan Rumus (6).

3.7 Implementasi dan Pengujian

Implementasi dan pengujian menjelaskan langkah-langkah yang dilakukan dalam proses implementasi dan pengujian metode yang digunakan pada penelitian. Proses ini mencakup penerapan algoritma K-Nearest Neighbors (KNN) berdasarkan parameter dan pendekatan yang telah ditentukan, serta evaluasi performa model menggunakan dataset yang tersedia.

3.7.1 Implementasi

Implementasi sistem adalah prosedur tersistematika yang dilakukan dalam menyelesaikan desain pada dokumen yang disetujui. Penelitian ini membutuhkan perangkat pendukung yaitu perangkat keras. Spesifikasi perangkat keras adalah sebagai berikut :

<i>Processor</i>	: Intel® Core™ i5-1035G1 CPU @ 1.00GHz (8 CPUs), ~1.2GHz
<i>RAM</i>	: 8,00 GB
<i>SSD</i>	: 256 GB

Spesifikasi perangkat lunak:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Platform : Google Collab

Browser : Google Chrome

Bahasa pemrograman : Python.

3.7.2 Pengujian

Pada tahapan ini dilakukan pengujian untuk metode penelitian yang digunakan, berikut pengujian yang dilakukan :

1. Pengujian Akurasi

Pengujian ini menggunakan *Confussion matrix* di mana salah satu cara yang sering digunakan padaproses evaluasi model *data mining* klasifikasi dengan memprediksi kebenaran objek. *Confusion Matrix* membantu dalam mengevaluasi kinerja model dan membuat analisis yang lebih menyeluruh daripada hanya menghitung probabilitas klasifikasi yang sebenarnya (akurasi). Rumus untuk menghitung akurasi dirujuk pada Rumus (7).

2. Pengujian K

Dalam penelitian ini, dilakukan pengujian terhadap berbagai nilai K pada algoritma *K-nearest neighbors* untuk menentukan nilai K optimal yang menghasilkan akurasi tertinggi. Dataset menggunakan Kaggle Dataset yang dimana beberapa atribut telah diseleksi oleh dokter spesialis saraf di pekanbaru, diperoleh data dan atribut sebanyak 15000 data, 14 atribut yang terdiri dari 13 atribut fitur dan 1 label kelas. Data ini dipisahkan menjadi 90% data latih dan 10% data uji, 80% data latih dan 20% data uji dan 70% data latih dan 30% data uji. Pengujian dilakukan dengan rentang nilai K dari 1 hingga 11 bilangan ganjil.

3.7.3 Evaluasi

Untuk memastikan bahwa model dapat dikembangkan dan digunakan untuk membuat keputusan, tahap evaluasi dilakukan untuk mengevaluasi teknik data mining atau model yang menghasilkan hasil yang baik. Dalam penelitian ini, Matrix Confusion adalah teknik yang sering digunakan untuk mengevaluasi model data mining klasifikasi dengan memprediksi kebenaran objek.

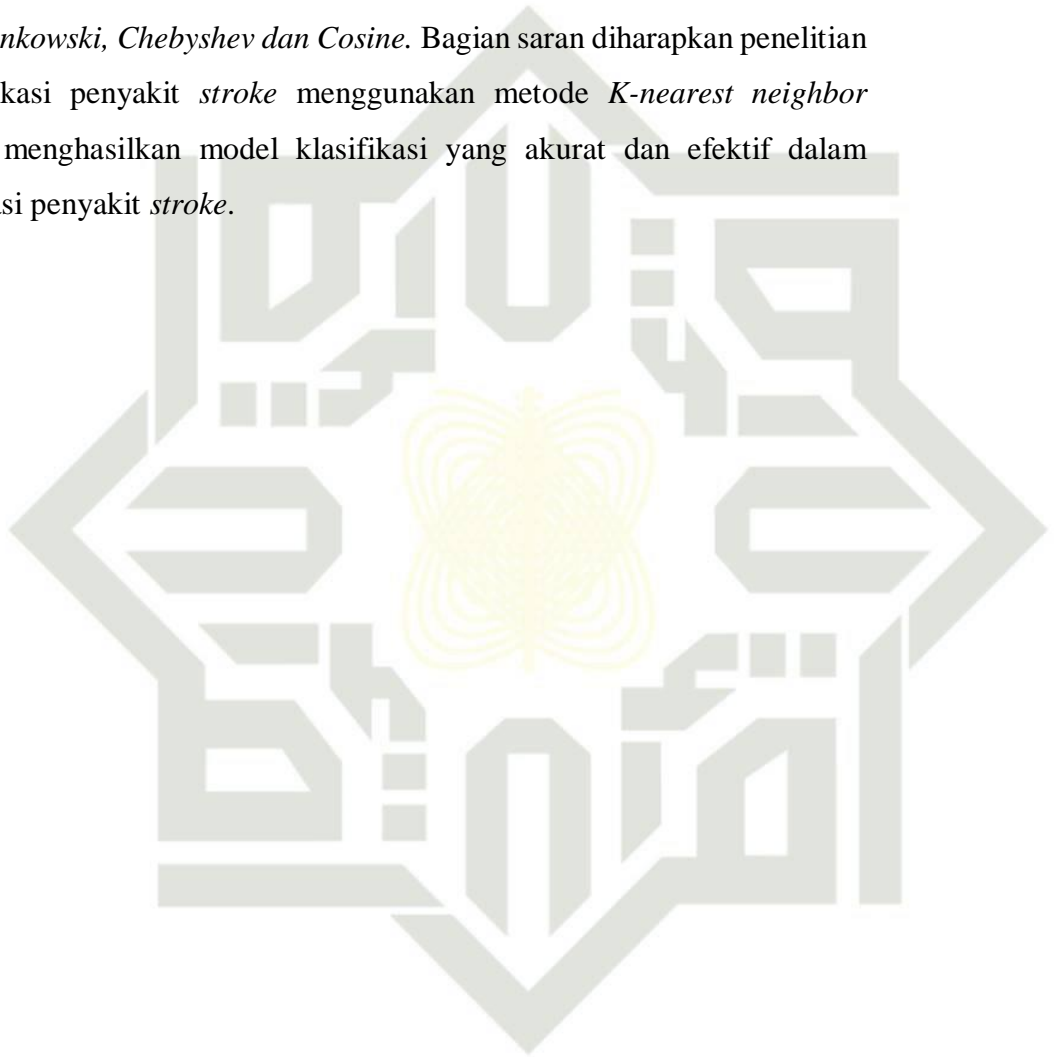


Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.8 Kesimpulan dan Saran

Tahapan kesimpulan ini adalah tahapan untuk menentukan hasil yang mana telah dilakukannya pengujian. Hal tersebut bertujuan untuk menentukan jumlah akurasi dari hasil yang telah kita dapatkan untuk menentukan apakah pasien terkena *stroke* atau tidak *stroke* dengan memperbandingkan perhitungan jarak *Euclidean*, *Manhattan*, *Minkowski*, *Chebyshev* dan *Cosine*. Bagian saran diharapkan penelitian tentang klasifikasi penyakit *stroke* menggunakan metode *K-nearest neighbor* (KNN) dapat menghasilkan model klasifikasi yang akurat dan efektif dalam mengidentifikasi penyakit *stroke*.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 5 PENUTUP

5.1 Kesimpulan

Dari hasil implementasi dan pengujian yang sudah dilakukan, dapat diambil kesimpulan sebagai berikut.

1. Klasifikasi Penyakit *Stroke* menggunakan metode *K-nearest neighbor* (KNN) dengan membandingkan perhitungan jarak (*Euclidean, Manhattan, Minkowski, Chebyshev dan Cosine*) sudah berhasil dilakukan.
2. Dari beberapa pengujian dihasilkan Akurasi tertinggi menggunakan pengujian semua fitur pada dataset penyakit *stroke*, dengan nilai akurasi 53.33%, presisi 53.33%, recall 53.33%, rasio perbandingan 90:10, $K=3$ dan menggunakan perhitungan jarak *euclidean*. Kategori akurasi pada dataset (Rendah) ini dikarenakan dari Berdasarkan visualisasi t-SNE, terdapat tumpang tindih yang signifikan antara kelas *stroke* dan *no stroke*. Tumpang tindih yang tinggi menunjukkan kesulitan model untuk membedakan kelas, dari metode klasifikasi yang digunakan.

5.2 Saran

1. Pada penelitian selanjutnya, diharapkan memakai metode optimasi atau melakukan penanganan data yang tidak seimbang untuk melihat perbandingan hasil akurasi yang lebih baik.
2. Pada penelitian selanjutnya, dapat menambahkan preprocessing data tambahan feature engineering membuat atribut baru yang lebih diskriminatif berdasarkan domain knowledge, atau mengurangi noise dalam data yang dapat memengaruhi kemampuan atribut untuk menjadi diskriminatif.
3. Pada penelitian selanjutnya, diharapkan menggunakan model lain seperti *Random Forest, Gradient Boosting (XGBoost, LightGBM), atau Neural Network* untuk mengatasi hubungan non-linier yang kompleks dalam data.

DAFTAR PUSTAKA

- Akmal, K., Faqih, A., & Dikananda, F. (2023). Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 470–477. <https://doi.org/10.36040/jati.v7i1.6367>
- Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. *SINTECH (Science and Information Technology) Journal*, 5(2), 191–197. <https://doi.org/10.31598/sintechjournal.v5i2.1222>
- Aziz, A., Pamungkas, D., & Setiawan, A. (2021). Analisa Perbandingan Algoritma Euclidean Dan Manhattan Distance Dalam Identifikasi Wajah. *Seminar Nasional Inovasi Teknologi*, 219–224.
- Ismafillah, D., Rohana, T., & Cahyana, Y. (2023). Implementasi Model Support Vector Machine dan Logistic Regression Untuk Memprediksi Penyakit Stroke. *JURIKOM (Jurnal Riset Komputer)*, 10(1), 2407–389. <https://doi.org/10.30865/jurikom.v10i1.5478>
- Juppreen, & Arnomo, S. A. (2021). Jurnal Comasie Jurnal Comasie. *PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PENCAPAIAN PROFIT PADA PT INDOLAND BATAM*, 5(2), 45–53. [http://ejournal.upbatam.ac.id/index.php/comasiejournal%0AJurnal Comasie ISSN \(Online\) 2715-6265%0APERANCANGAN](http://ejournal.upbatam.ac.id/index.php/comasiejournal%0AJurnal%20Comasie%20ISSN%20(Online)%202715-6265%0APERANCANGAN)
- Martono, G. H., & Sulistianingsih, N. (2024). Perbandingan Matriks Jarak Pada Algoritma K-NN Untuk Prediksi Penyakit Diabetes. *JoMI: Journal of Millennial Informatics*, 2(1), 1–6.
- Mujammil, A., Oktavianto, H., & Abdurrahman, G. (2021). *KLASIFIKASI PENYAKIT STROKE MENGGUNAKAN ALGORITMA KNN DENGAN OPTIMASI TEKNIK BAGGING. 1610651042.*
- Zakmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2).

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>

Pinzon, R. T. (n.d.). *Awat Stroke* (Issue 112).

Prasath, V. B. S., Arafat, H., Alfeilat, A., Hassanat, A. B. A., Lasassmeh, O., & Ahmad, S. (n.d.). *Effects of Distance Measure Choice on KNN Classifier Performance - A Review*. 1–39.

Rahayu, S., & Yamasari, Y. (2024). Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM). *Journal of Informatics and Computer Science*, 05, 440–446.

Rahel Lina Simanjuntak, Rizki Agung Ramadhan, Theresia Romauli Siagian, & Vina Anggriani. (2023). Komparasi Algoritma KNN dan SVM dalam Memprediksi Penyakit Stroke. *Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 3(3), 60–74. <https://doi.org/10.55606/teknik.v3i3.2474>

Suputri, N. D., Khalid, K., Rolliawati, D., Informasi, P. S., Islam, U., Sunan, N., Gresik, K., Timur, J., Disease, D. S., & Tree, D. (2022). *Komparasi Penerapan Metode Bagging dan Adaboost pada Algoritma C4. 5 untuk Prediksi Penyakit Stroke*. 11(September), 567–577.

Setiawan, A. (2022). Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris. *Jurnal Sains Dan Edukasi Sains*, 5(1), 28–37. <https://doi.org/10.24246/juses.v5i1p28-37>

Setiawan, A., Waleska, R. F., Purnama, M. A., & Efrizoni, L. (2024). Komparasi Algoritma K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), dan Decision Tree dalam Klasifikasi Penyakit Stroke. *Jurnal Informatika &*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Rekayasa Elektronika, 7(1), 107–114.

Stiawan et al. (2021). Diagnosis Dan Tatalaksana Stroke Hemoragik. *Jurnal Medika Utama*, 02(01), 402–406.

Slaeman, R. (2022). Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke Support Vector Machine Algorithm Analysis In Stroke Disease Classification. 9(3), 922–928.

Uni Umamatun Nysa, J., Mahmudi, A., & Auliasari, K. (2023). Perbandingan Jarak Euclidean, Manhattan, Chebyshev Pada Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbors (Knn). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2443–2450. <https://doi.org/10.36040/jati.v7i4.7544>

Utama, Y. A., & Nainggolan, S. S. (2022). Faktor Resiko yang Mempengaruhi Kejadian Stroke: Sebuah Tinjauan Sistematis. *Jurnal Ilmiah Universitas Batanghari Jambi*, 22(1), 549. <https://doi.org/10.33087/jiubj.v22i1.1950>

Wahyono, W., Trisna, I. N. P., Sariwening, S. L., Fajar, M., & Wijayanto, D. (2020). Comparison of distance measurement on k-nearest neighbour in textual data classification. *Jurnal Teknologi Dan Sistem Komputer*, 8(1), 54–58. <https://doi.org/10.14710/jtsiskom.8.1.2020.54-58>

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>

Wijaya, N., Endah, M., & Feliati, M. (2020). Penerapan Algoritma Decision Tree C.45 Untuk Klasifikasi Data Status Huni Rumah Rehabilitasi Pasca Erupsi Merapi. 424–430.

Zriati, Z., & Qomariyah, N. (2022). Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN). *ROUTERS: Jurnal Sistem Dan Teknologi Informasi*, 1(1), 1–8. <https://doi.org/10.25181/rt.v1i1.2665>

DAFTAR RIWAYAT HIDUP



Informasi Personal

Nama : SHAFIQ ZAHIDAN
 Tempat/Tanggal Lahir : Pekanbaru, 4 Juli 2001
 Alamat : Jl. Manyar Sakti Gg. Kayu Manis
 Jenis Kelamin : Laki-laki
 Agama : Islam
 Tinggi badan : 170 cm
 Kewarganegaraan : Indonesia
 Hobby : Olahraga, nonton, game
 Motto : “Percaya dan yakin dengan potensi diri”

Riwayat Pendidikan

Tahun 2006-2007 : TK Babussalam
 Tahun 2007-2013 : SD Babussalam
 Tahun 2013-2017 : SMPIT Imam An-Nawawi
 Tahun 2017-2020 : MAN 2 Pekanbaru

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.