

**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.

**PENERAPAN METODE *NAÏVE BAYES CLASSIFIER* UNTUK
MENGKLASIFIKASI PENYAKIT *STROKE*****TUGAS AKHIR**

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

Oleh

RAYHAN ATHALLAH IRSAN

NIM. 12050116470



UIN SUSKA RIAU

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

LEMBAR PERSETUJUAN

**PENERAPAN METODE *NAÏVE BAYES CLASSIFIER* UNTUK
MENGKLASIFIKASI PENYAKIT *STROKE***

TUGAS AKHIR

Oleh

RAYHAN ATHALLAH IRSAN**NIM. 12050116470**

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 6 Januari 2025

Pembimbing,

SISKA KURNIA GUSTI, S.T., M.Sc.**NIP. 198610092022032001**



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

LEMBAR PENGESAHAN

**PENERAPAN METODE *NAÏVE BAYES CLASSIFIER* UNTUK
MENGKLASIFIKASI PENYAKIT *STROKE***

Oleh

RAYHAN ATHALLAH IRSAN

NIM. 12050116470

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, 6 Januari 2025

Mengesahkan,
Ketua Jurusan,

DR. HARTONO, M.Pd.

NIP. 19640301 199203 1 003

IWAN ISKANDAR, M.T.

NIP. 19821216 201503 1 003

DEWAN PENGUJI

- | | |
|--------------|-----------------------------------|
| Ketua | : Iwan Iskandar, M.T. |
| Pembimbing I | : Siska Kurnia Gusti, S.T., M.Sc. |
| Penguji I | : Elvia Budianita, ST, M.Cs |
| Penguji II | : Iis Afrianty, S. T., M. Sc |

**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 : Rayhan Athallah Irsan
NIM : 12050116470
Tempat/Tgl.Lahir : Pekanbaru/11 Desember 2001
Fakultas : Sains dan Teknologi
Prodi : Teknik Informatika
Judul Skripsi : PENERAPAN METODE NAÏVE BAYES CLASSIFIER
UNTUK MENGLASIFIKASI 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, 6 Januari 2025
Yang membuat pernyataan



Rayhan Athallah Irsan
NIM. 12050116470



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.



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 terdapat dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 6 Januari 2025

Yang membuat pernyataan,

RAYHAN ATHALLAH IRSAN
12050116470

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.



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 Bunda 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 Bunda 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 darimua 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

Penyakit *stroke* merupakan masalah yang serius pada negara dengan penghasilan rendah dibandingkan negara berpenghasilan tinggi. Menurut *World Stroke Organization* (2022), lebih dari 12,2 juta atau satu dari empat orang di atas usia 25 mengalami *stroke* atau lebih dari 101 juta orang yang hidup saat ini. Sehingga diterapkanlah teknik *data mining* menggunakan algoritma *naïve bayes classifier* untuk mengklasifikasi penyakit *stroke*. Data yang digunakan diambil dari *kaggle* dengan jumlah data sebanyak 15000 data pasien *stroke*. didapatkan hasil akurasi tertinggi menggunakan *Feature Selection* dengan jumlah data sebanyak 12500 data dengan 9 atribut sebesar 53 %, presisi 53%, dan nilai recall 53%.

Kata kunci : *stroke*, *data mining*, *naïve bayes classifier*, klasifikasi

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 serious problem in low-income countries compared to high-income countries. According to the World Stroke Organization (2022), more than 12.2 million or one in four people over the age of 25 have a stroke or more than 101 million people alive today. So that the data mining technique is applied using the naïve bayes classifier algorithm to classify stroke. The data used was taken from kaggle with a total of 15,000 stroke patient data. The highest accuracy results were obtained using Feature Selection with a total of 12,500 data with 9 attributes of 53%, precision of 53%, and recall value of 53%.

Keywords: classification, data mining, naïve bayes classifier, 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.

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 Surya Agustian, S.T., 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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6. Ibu Elvia Budianita, S.T., M.Cs., 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 tuaku tersayang, yang selalu memberi dukungan, semangat dan mendoakan tanpa henti hingga selesainya tugas akhir ini.
9. Adik – adikku tersayang, yang selalu menghibur dan menyemangati penulis dalam menyelesaikan laporan tugas akhir ini.
10. Anak – anak Skuy, teman seperjuangan dari Maba yang selalu ada dan sangat membantu penulis dalam menyelesaikan Tugas Akhir ini.
11. Anak – anak Warung Bude, teman tongkrongan yang ku kenal jauh sebelum masuk dunia perkuliahan yang selalu ada disaat susah atau pun senang dan sangat membantu penulis dalam menyelesaikan Tugas Akhir ini
12. Semua teman-teman yang sudah membantu, menyemangati, dan menghibur sehingga penulis dapat menyelesaikan laporan tugas Akhir ini.
13. Seluruh pihak yang belum kami cantumkan, terima kasih atas dukungannya, baik material maupun spiritual.
14. Untuk diri sendiri, Terimakasih telah berhasil melangkah sejauh ini dengan banyaknya cobaan yang ada. Terimakasih sudah memberikan dirimu lebih berkembang hampir disetiap harinya baik itu fisik, pola pikir, dan mentality. Terimakasih atas beberapa hal yang membuatmu bangga dengan dirimu sendiri yang mungkin dinilai orang lain sebagai hal yang kecil. Terimakasih untuk selalu berusaha membuat dirimu bahagia, senang dan tertawa dengan orang - orang terdekat yang kamu senangi. Terimakasih karena selalu bersyukur atas segala momen yang berharga dihidupmu. Tetaplah rendah hati kepada orang lain dan jangan lupa selalu cintai dirimu terlebih dahulu sebelum kamu mencintai orang lain.

Semikian 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

**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 mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

membangun sangat kami harapkan untuk kesempurnaan laporan ini dapat disampaikan melalui email 12050116470@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, 6 Januari 2025

Penulis



UIN SUSKA RIAU



DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN.....	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN.....	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR.....	ix
DAFTAR ISI	xii
DAFTAR GAMBAR	xv
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 <i>Data Mining</i>	5
2.1.1 Tahapan <i>Knowledge Discovery in Database</i>	5
2.2 Klasifikasi	7
2.3 <i>Naïve Bayes Classifier</i>	7

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.

© Hak cipta milik UIN Suska Riau	2.3.1	Pemodelan Algoritma <i>Naïve Bayes</i>	8
	2.4	<i>Stroke</i>	9
	2.5	<i>Confusion Matrix</i>	12
	2.6	Penelitian Terkait.....	13
	BAB 3 METODOLOGI PENELITIAN.....		21
	3.1	Perumusan Masalah.....	22
	3.2	Pengumpulan Data.....	22
	3.3	Analisa Data.....	22
	3.3.1	Analisa Kebutuhan Data.....	23
	3.3.2	<i>Naïve Bayes Classifier</i>	26
	3.4	Implementasi dan Pengujian.....	27
	3.4.1	Implementasi	28
	3.4.2	Pengujian	28
	3.5	Kesimpulan dan Saran	28
	BAB 4 PEMBAHASAN		29
State Islamic University of Sultan Syarif Kasim Riau	4.1	Analisa Data.....	29
	4.2	Analisa Kebutuhan Data.....	29
	4.2.1	Pengumpulan Data	29
	4.2.2	Selection Data.....	30
	4.2.3	Cleaning Data.....	32
	4.2.4	Transformasi Data.....	33
	4.3	<i>Naïve Bayes Classifier</i>	35
	4.4	Implementasi.....	44
	4.4.1	Batasan Implementasi.....	44
	4.4.2	Lingkungan Implementasi	44

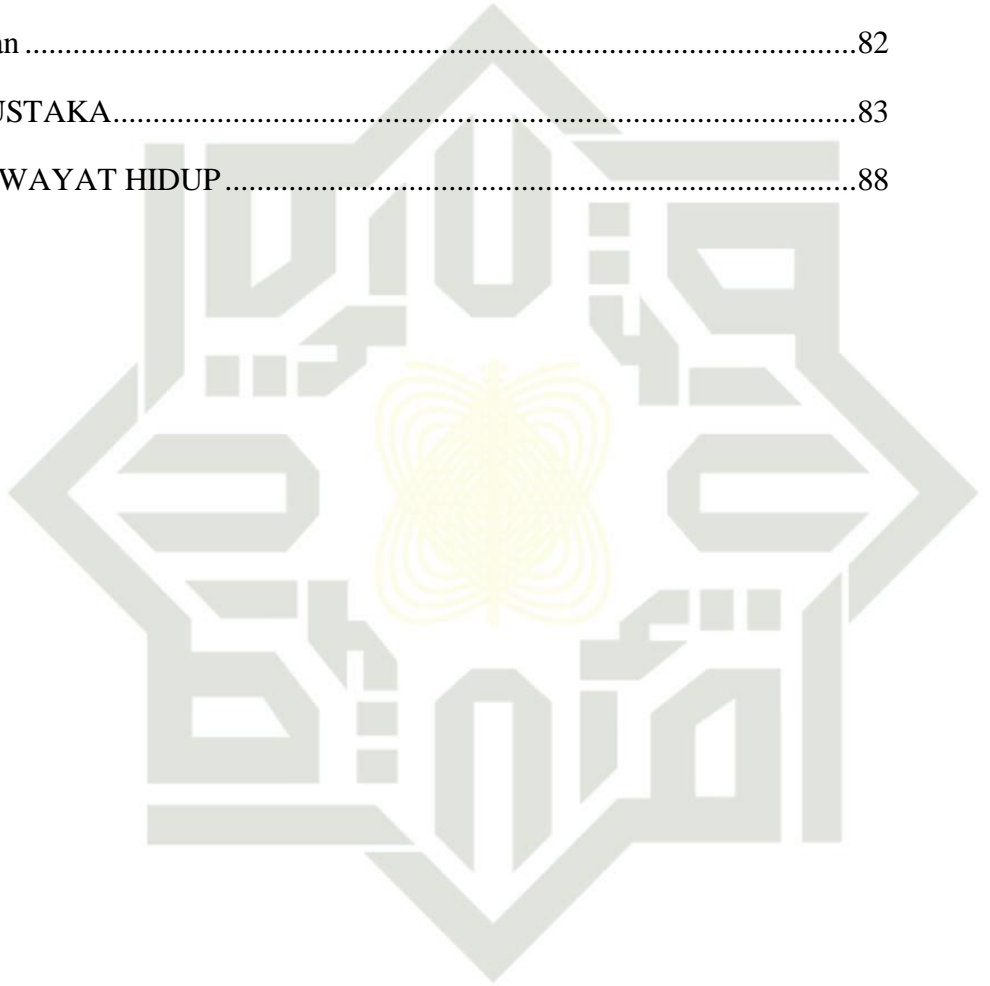


- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. 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 milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

4.4.3	Tahapan Implementasi.....	45
4.5	Evaluasi.....	47
4.5.1	Kesimpulan Pengujian.....	49
BAB 5 PENUTUP.....		82
5.1	Kesimpulan.....	82
5.2	Saran.....	82
DAFTAR PUSTAKA.....		83
DAFTAR RIWAYAT HIDUP.....		88



UIN SUSKA RIAU



DAFTAR GAMBAR

Gambar 1 Tahapan Knowledge Discovery in Database	6
Gambar 2. Tahapan Metodologi Penelitian	21
Gambar 3. Flowchart Naïve Bayes Classifier	26
Gambar 4. Import Modul	45
Gambar 5. Mount Google Drive.....	45
Gambar 6. Encode Data.....	45
Gambar 7. Memisahkan Dataset Fitur Dan Label	45
Gambar 8. Split Data Train Dan Test	46
Gambar 9. Inisialisasi Model Naive Bayes.....	46
Gambar 10. Rasio Perbandingan 90:10.....	47
Gambar 11. Rasio Perbandingan 80:20.....	48
Gambar 12. Rasio Perbandingan 70:30.....	49

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 TABEL

Tabel 1. Confusion Matrix	12
Tabel 2. Penelitian Terkait	13
Tabel 3. Dataset Stroke	22
Tabel 4. Transformasi Data	24
Tabel 5. Dataset Penyakit Stroke.....	29
Tabel 6. Deskripsi Dataset Awal	30
Tabel 7. Hasil Transformasi Data.....	33
Tabel 8. Nilai Probabilitas Fitur Gender	37
Tabel 9. Nilai Probabilitas Fitur Hypertension.....	37
Tabel 10. Nilai Probabilitas Fitur Heart Disease	38
Tabel 11. Nilai Probabilitas Fitur Average Glucose Level.....	38
Tabel 12. Nilai Probabilitas Fitur Body Mass Index (BMI).....	38
Tabel 13. Nilai Probabilitas Fitur Smoking Status	39
Tabel 14. Nilai Probabilitas Fitur Alcohol Intake	39
Tabel 15. Nilai Probabilitas Fitur Dietary Habits.....	39
Tabel 16. Nilai Probabilitas Fitur HDL.....	40
Tabel 17. Nilai Probabilitas Fitur LDL	40
Tabel 18. Nilai Probabilitas Fitur Systolic	40
Tabel 19. Nilai Probabilitas Fitur Diastolic.....	41
Tabel 20. Hasil Pengujian Perbandingan Rasio.....	50

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

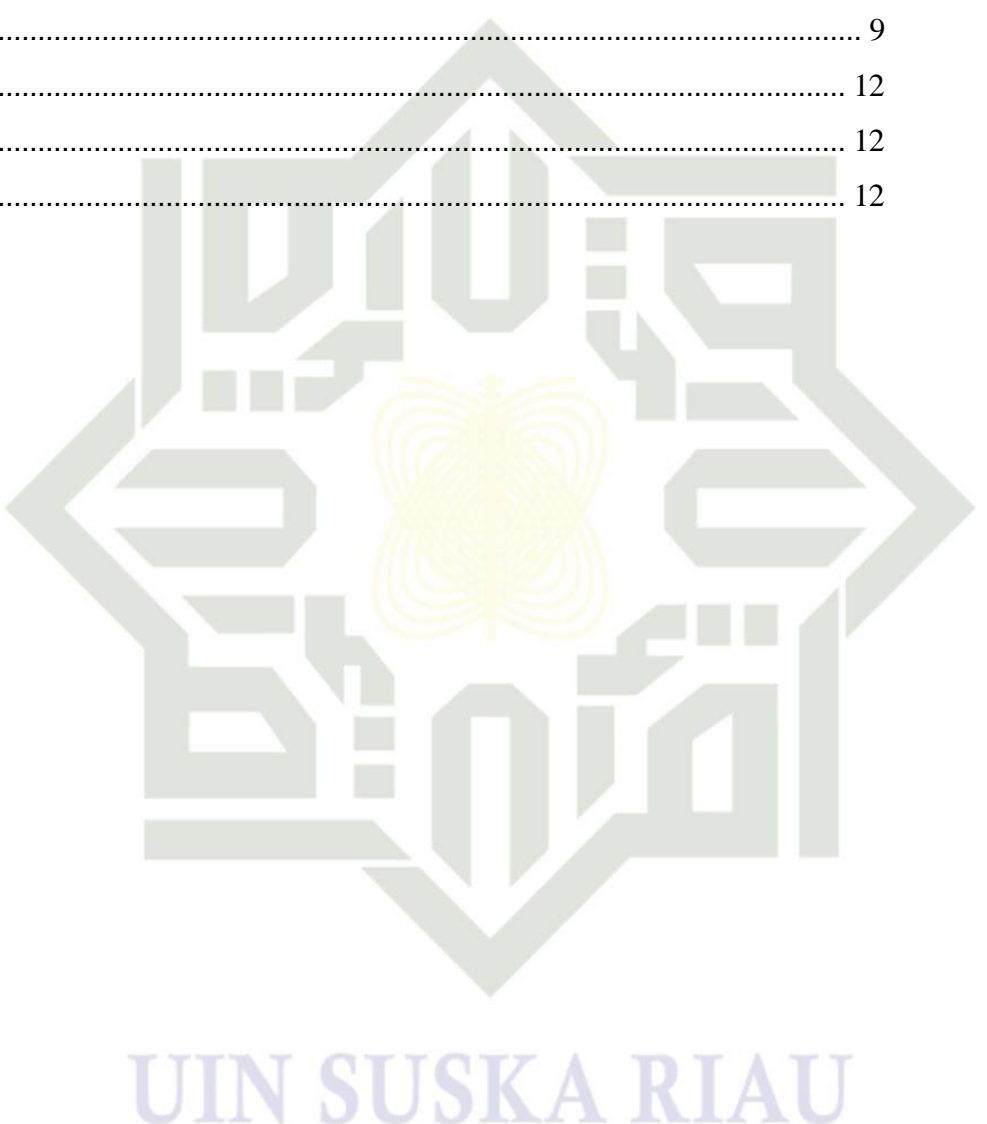
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

DAFTAR RUMUS

© Hak cipta	8
Ilmu	8
Keislaman	9
Keindonesiaan	9
Keberagaman	12
Keberadaban	12
Keberlanjutan	12
Keberkeadilan	12

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

Disejumlah negara dengan penghasilan rendah, penyakit *stroke* merupakan isu serius pada bidang kesehatan dibanding dengan negara berpenghasilan tinggi (Akmal et al., 2023). Sistem saraf pusat berfungsi dengan tidak normal, yang menyebabkan gangguan aliran darah ke otak sehingga 15 juta kasus *stroke* terjadi setiap tahun (Gustin Rahayu, 2023). Penyakit *stroke* dapat menyebabkan penurunan atau bahkan menghilangkan fungsi pada jaringan. Salah satu gejalanya seperti kelemahan otot anggota gerak tubuh, seperti jari-jari tangan. Otot bagian tangan ini sangat penting untuk melakukan aktivitas sehari-hari. Kelemahan otot membuat sebagian orang bergantung pada orang lain. Rehabilitasi adalah cara terbaik untuk mengurangi risiko kecacatan setelah serangan *stroke* (Putra Kusuma et al., 2022).

Penyebab dari penyakit *stroke* itu terjadi dikarenakan berbagai faktor seperti : tekanan darah tinggi, riwayat *fibrilasi atrium*, *kolesterol*, *diabetes*, dan faktor lain. Penanganan penyakit *stroke* dilakukan secara manual yang diperiksa oleh dokter penyakit syaraf kemudian didiagnosa dengan mengajukan pertanyaan tentang keluhan pasien dan potensi penyebab *stroke*. Dengan demikian, pasien dapat menentukan tingkat risiko terkena *stroke* (Puspitawuri et al., 2019). Berdasarkan laporan dari *World Stroke Organization* (2022), bahwa lebih dari 12,2 juta orang setidaknya satu dari empat orang di atas usia 25 tahun terkena gejala *stroke* atau lebih dari 101 juta orang yang hidup saat ini akan mengalami *stroke* dan 7,6 juta atau 62% dari jenis *stroke iskemik* terjadi setiap tahun. Perdarahan *intracerebral* yang mencakup 1,2 juta atau lebih dari 28% perdarahan *subarachnoid* terjadi pada semua kejadian *stroke*. Penderita penyakit *stroke* baru di Amerika Serikat sekitar 795 ribu pasien dan berulang sekitar 610 ribu di antaranya adalah *stroke* pertama (Dwilaksono et al., 2023).

Teknik klasifikasi menggunakan *data mining* untuk penyakit *stroke* merupakan teknik untuk mengidentifikasi pola, tren, dan pengetahuan terbaru dari



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

Jumlah data besar yang tersimpan dalam *repositori* atau tempat penyimpanan dengan menggunakan cara pengenalan pola statistik dan matematika (Abidin et al., 2023). Proses yang melibatkan pengumpulan dan penggunaan data historis atau bisa disebut *Knowledge Discovery in Database (KDD)* merupakan proses untuk menemukan keteraturan, pola, atau hubungan dalam kumpulan data yang besar. Adapun tahapan dalam menganalisis *data mining* dengan pendekatan KDD seperti *selection data, preprocessing, transformasi, data mining*, dan evaluasi (Fadri, 2023).

Naïve Bayes Classifier adalah metode klasifikasi statistik yang digunakan untuk memprediksi kemungkinan dari anggota kelas. Metode *Naïve Bayes Classifier* memberikan tingkat Akurasi dan kecepatan yang sangat tinggi ketika digunakan dalam aplikasi dengan data yang besar (Titimeidara & Hadikurniawati, 2021). Menurut penelitian yang dilakukan oleh (Riany & Testiana, 2023b) Faktor – faktor penyebab stroke diantaranya jenis kelamin, usia, riwayat penyakit hipertensi, riwayat penyakit jantung, status pernikahan, jenis pekerjaan, tipe tempat tinggal, kadar glukosa, index masa tubuh dan status merokok. Algoritma *Naïve Bayes* dapat digunakan untuk mengklasifikasikan data penyakit *stroke* dengan hasil akurasi 92,48%.

Penelitian yang dilakukan oleh (Rahma et al., 2023) untuk mendiagnosis *stroke* dini. Data *stroke* yang digunakan berasal dari Rumah Sakit Haji Surabaya, yang berjumlah 100 data dengan 10 variabel dan 2 kelas klasifikasi, yaitu *stroke* dan tidak *stroke*. Penelitian ini menggunakan metode *Naïve Bayes* yang menghasilkan *10-Folds Cross Validation* dengan hasil *oversampling* sebesar 71% dan *undersampling* sebesar 64%. Kesimpulannya bahwa *Naïve Bayes* standar lebih baik dibandingkan dengan *Naïve Bayes* dengan penanganan data tidak seimbang.

Penelitian terkait mengenai klasifikasi metode *Naïve Bayes* oleh (Irmayani, 2021), Proses klasifikasi digunakan untuk memprediksi hewan yang ada pada kebun binatang. Terdapat 7 label binatang yang akan diklasifikasi menggunakan aplikasi *Orange* diantaranya: *amphibian, bird, fish, insect, invertebrate, mammal* dan *reptile* pada kebun binatang. *Proportion of prediction* menunjukkan ada 4



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

prediksi klasifikasi dengan nilai 100% : hasil prediksi pertama menghasilkan 92,9%, hasil prediksi kedua menghasilkan 80%, hasil prediksi ketiga menghasilkan 66,7% dan hasil prediksi terakhir menghasilkan 33,3%. Hasil *visualisasi* untuk klasifikasi *Naive Bayes* digambarkan dengan grafik garis pada *Receiver Operating Characteristics Analysis*.

Peneliti menyimpulkan bahwa akan melakukan penelitian tugas akhir yang berjudul “*Penerapan Metode Naïve Bayes Classifier Untuk Mengklasifikasi Penyakit Stroke*”. Penelitian ini diharapkan dapat mengetahui dan memprediksi penyakit *stroke* dengan akurasi yang diharapkan.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian untuk laporan tugas akhir berdasarkan latar belakang yang sudah dijelaskan yaitu :

1. Bagaimana penerapan metode *Naïve Bayes Classifier* untuk mengklasifikasikan penyakit *stroke*
2. Bagaimana hasil akurasi tertinggi dari perbandingan evaluasi beberapa pengujian pada klasifikasi penyakit *stroke* menggunakan metode *Naïve Bayes Classifier*.

1.3 Batasan Masalah

Penetapan luas dari dari suatu masalah pada penelitian diperlukan batasan masalah. Peneliti membatasi cakupan topik yang akan dibahas adalah:

1. *Dataset* yang digunakan pada penelitian yaitu *dataset* pasien *stroke* sebanyak 15000 data yang diambil dari situs *kaggle*. Atribut *fitur* dan *label* yang digunakan berdasarkan hasil wawancara dengan dokter spesialis saraf RS Awal Bross Pekanbaru terdiri dari 13 atribut fitur dan 1 label kelas.
2. Atribut yang digunakan sesuai dengan hasil wawancara dengan dokter spesialis saraf RS Awal Bross Pekanbaru yaitu *Age*, *Gender*, *Hypertension*, *Heart Disease*, *Average Glucose Level*, *Body Mass Index*



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.

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

3. Pengklasifikasian pasien penyakit *stroke* yang akan diteliti dengan kelas *stroke* dan *no stroke*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagaimana peneliti menjabarkan sebagai berikut:

1. Penerapan metode *Naïve Bayes Classifier* dalam mengklasifikasikan penyakit *stroke*.
2. Memperoleh hasil akurasi tertinggi dari perbandingan evaluasi beberapa pengujian pada klasifikasi penyakit *stroke* menggunakan metode *Naïve Bayes Classifier*.

1.5 Manfaat Penelitian

Manfaat penelitian penyakit *stroke* menggunakan metode *Naïve Bayes Classifier* dalam klasifikasi *stroke* antara lain :

1. Menambah pengalaman dan wawasan dalam melakukan penelitian untuk penerapan metode *Naïve Bayes Classifier*.
2. Mengetahui performa dari metode *Naïve Bayes Classifier* dalam melakukan klasifikasi terhadap 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 mengemukakan 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 merupakan proses mengekstraksi suatu pengetahuan penting dari ‘data banyak’ melalui penggunaan teknik komputasi dan analisis statistik. Proses *data mining* melibatkan pemilihan data, *preprocessing* data, pemodelan data, evaluasi, dan interpretasi untuk mengidentifikasi pola, tren, prediksi, dan pengambilan keputusan yang lebih baik (Jane Indah Pratiwi et al., 2022). Proses *data mining* sudah digunakan dalam berbagai bidang untuk mendapatkan informasi yang mungkin tersembunyi dalam data seperti bisnis, sains, kesehatan, dan masih banyak lagi (Sulianta, 2023).

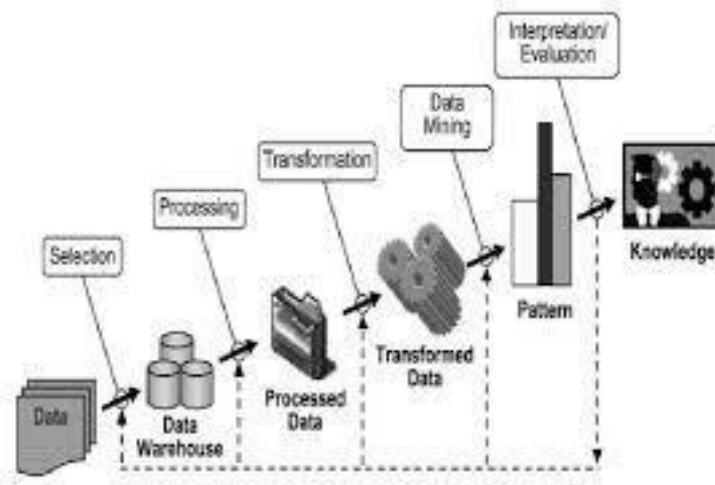
Data mining adalah proses basis data untuk menemukan pola - pola tersembunyi, informasi, dan prediksi yang terabaikan oleh para pelaku bisnis. Sudah banyak perusahaan – perusahaan saat ini yang menggunakan *data mining* untuk mencari permasalahan bisnis (Journal et al., 2022). Segala sesuatunya dapat berkembang pesat dikarenakan era data yang semakin banyak tersedia dan berkembang (Sianturi et al., 2019).

2.1.1 Tahapan *Knowledge Discovery in Database*

Tahapan untuk memperoleh pengetahuan dari database atau disebut *Knowledge Discovery In Database* (KDD) yang berisikan tabel - tabel yang saling berhubungan / berelasi. Proses KDD dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan pengambilan keputusan. *Knowledge Discovery in Database* (KDD) dan *data mining* sering digunakan secara bergantian dalam proses penggalian informasi tersembunyi pada suatu basis data yang besar (Alghifari & Juardi, 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.



Gambar 1 Tahapan Knowledge Discovery in Database

Berdasarkan penelitian yang ada, proses KDD dapat digambarkan sebagai berikut (Yuli Mardi, 2019) :

1. Data Selection

Pemilihan data dilakukan sebagai tahapan awal untuk penggalian informasi dari sekumpulan data yang diambil untuk digunakan dalam proses *data mining* yang akan disimpan dalam berkas berbeda dari basis data operasional.

2. Pre-processing / Cleaning

Pre-processing / cleaning data dilakukan untuk membuang duplikasi data, memeriksa data yang *inconsisten*, dan memperbaiki kesalahan pada data. Proses *cleaning* pada data menjadi fokus utama pada *Knowledge Discovery in Database* (KDD) seperti halnya proses *enrichment* yang dapat membuat data yang sudah ada menjadi data yang lebih relevan untuk digunakan.

3. Transformation

Transformation data digunakan agar pemrosesan coding menjadi lebih kreatif dengan jenis atau pola yang akan dicari dalam basis data.

4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD).



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

Interpretation / Evaluation

Tahap *evaluation* merupakan bagian dari proses KDD yang mencakup pemeriksaan pola atau informasi yang ditemukan bertentangan dengan hipotesis atau fakta sebelumnya kedalam bentuk yang mudah dipahami oleh semua pihak yang terlibat.

2.2 Klasifikasi

Klasifikasi adalah proses untuk menemukan sekumpulan model yang dapat menjelaskan kelas – kelas data untuk memprediksi kelas yang belum diketahui dari suatu objek. Proses klasifikasi melibatkan analisis terhadap data latih (*training set*) untuk membangun model, dan data uji (*test set*) untuk mengevaluasi akurasi model. Proses klasifikasi dapat digunakan untuk memprediksi nama atau nilai dari objek data berdasarkan karakteristik yang ada (Sari & M, 2024).

Klasifikasi merupakan salah satu teknik untuk penggalian data yang melibatkan pengelompokan data ke dalam kategori atau kelas yang telah ditentukan berdasarkan atribut yang ada dan menghasilkan sebuah model yang dapat memprediksi kategori atau label data secara akurat berdasarkan pola yang tepat. (Rahayu et al., 2024).

2.3 Naïve Bayes Classifier

Naïve Bayes Classifier merupakan pengklasifikasian dengan cara menghitung sekumpulan probabilitas untuk jumlah frekuensi dan kombinasi nilai dari dataset. *Teorema bayes* merupakan algoritma yang dapat mengambil semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai variabel kelas (Nuralia et al., 2023).

Metode *naïve bayes classifier* didasarkan pada *teorema bayes* dan dikombinasikan dengan "*naïve*" yang berarti bahwa setiap atribut atau variabel bersifat bebas (*independent*). Metode ini dapat dilatih dengan efisien untuk suatu pembelajaran terawasi (*supervised learning*). Keuntungan dari klasifikasi yaitu dengan jumlah data pelatihan yang cukup kecil dapat memperkirakan parameter yang diperlukan seperti sarana dan varian dari variabel. Variabel *independent*



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

diasumsikan sebagai variasi dari variabel untuk masing – masing kelas tanpa menentukan keseluruhan *matriks kovarians*(Caron & Markusen, 2016).

Teorema perhitungan *naïve bayes classifier* berdasarkan probabilitas sebagai berikut :

$$P(C | F) = \frac{P(C).P(F | C)}{P(F)} = \dots\dots\dots (1)$$

Keterangan:

- P(C | F) : Probabilitas akhir bersyarat (C) terjadi jika diberikan bukti (F).
- P(F | C) : Probabilitas sebuah bukti (F) terjadi akan mempengaruhi kelas (C).
- P(C) : Probabilitas awal (C) terjadi tanpa memandang bukti apapun.
- P(F) : Probabilitas awal (E) terjadi tanpa memandang kelas / bukti yang lain.

2.3.1 Pemodelan Algoritma Naïve Bayes

Pemrosesan data dilakukan setelah data dikumpulan dan telah melalui *preprocessing* data. Proses pemodelan dengan algoritma *Naïve Bayes* sebagai berikut.

1. Nilai Probabilitas Prior

Probabilitas kelas dihitung dengan membagi jumlah data yang memiliki kelas dengan jumlah data pelatihan. Dibawah ini rumus untuk menghitung nilai probailitas kelas.

$$P(Y) = \frac{\text{jumlah sampel dengan kelas } Y}{N} = \dots\dots\dots (2)$$

Keterangan :

- P(Y) : Probabilitas prior dari kelas Y
- Jumlah sampel dengan kelas Y : Jumlah kemunculan kelas Y dalam data pelatihan.
- N : Total jumlah data dalam data pelatihan.

2. Nilai Probabilitas setiap Atribut

Probabilitas bahwa atribut X_i memiliki nilai tertentu jika kelas targetnya adalah Y. Probabilitas atribut X_i terhadap kelas Y dihitung dengan membagi jumlah data yang memiliki atribut X_i dan kelas Y dengan total jumlah data yang

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

memiliki kelas Y. Ini memberikan pemahaman tentang seberapa sering nilai atribut X_i muncul dalam kelas Y. Adapun rumus untuk mencari nilai probabilitas setiap fitur yaitu:

$$P(X = x|Y) = \frac{\text{Jumlah sampel dengan atribut } X=x \text{ dan Kelas } Y}{\text{Jumlah sampel pada kelas } Y} = \dots\dots\dots (3)$$

Keterangan:

$P(X=x|Y)$: Probabilitas bahwa atribut X_i memiliki nilai tertentu jika kelas targetnya Y.

Jumlah sampel dengan atribut $X=x$ dan Kelas Y : Jumlah data yang memiliki atribut X_i dan kelas Y.

Jumlah sampel pada kelas Y : Jumlah total data yang memiliki kelas Y

3. Nilai Prediksi

Rumus yang digunakan untuk menghitung nilai prediksi kelas target dari data uji berdasarkan probabilitas fitur yang dihitung sebelumnya yaitu:

$$P(Y|X) \propto P(Y) \times \prod_{i=1}^n P(X_i|Y) = \dots\dots\dots (4)$$

Keterangan :

$P(Y|X)$: Probabilitas kelas Y given data X, ini adalah apa yang ingin diprediksi atau ketahui setelah memiliki data X.

\propto : Proporsional

$P(Y)$: Probabilitas kelas Y

$\prod_{i=1}^n P(X_i|Y)$: Hasil perkalian dari probabilitas kondisional dari setiap atribut X_i given kelas Y, untuk semua atribut dalam data X.

4. **Stroke**

Penyebab penyakit *stroke* ditandai oleh gangguan fungsi otak fokal atau global yang berdampak pada kematian tanpa penyebab lain selain *vasikuler* (Putra Kusuma et al., 2022). *Stroke* dapat mengakibatkan penderitanya menjadi cacat dan kehilangan kemampuan untuk hidup mandiri serta secara signifikan hal ini dapat mempengaruhi ADL, kualitas hidup, psikologis, dan kesehatan sosial. *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.

merupakan penyebab kecacatan pertama dan kematian ketiga terbesar setelah penyakit *jantung koroner* dan *kanker*. *Stroke* dikategorikan menjadi 2 yaitu *stroke iskemik* dan *stroke hemoragik* (Dwilaksono et al., 2023).

Stroke iskemik terjadi dikarenakan penyumbatan pada aliran darah akibat dari trombosis atau emboli, sedangkan *stroke hemoragik* pada saat pembuluh darah yang pecah dan mengakibatkan perdarahan. *Stroke hemoragik* memiliki risiko morbiditas dan mortalitas yang lebih rendah dibanding dengan *stroke iskemik*. Arteri *servikal* atau *serebral* tersumbat yang menyebabkan jaringan otak mati merupakan faktor terjadinya *stroke iskemik*. Dua jenis *stroke hemoragik* terjadi ketika pembuluh darah *intracerebrum* pecah ke dalam ruang *subarachnoid*. Perdarahan *intracerebral* (ICS) merupakan perdarahan yang pada pembuluh darah bagian *parenkim* otak sedangkan perdarahan *subarachnoid* (SAH) merupakan perdarahan akut di luar pembuluh darah otak yang mengakibatkan pembuluh darah pada otak pecah (Haiga et al., 2022).

2.4.1.1 Faktor Risiko Stroke

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

1. Umur

Insidens stroke meningkat seiring dengan umur, risiko *stroke* meningkat dua kali lipat tiap dekade setelah umur 55 tahun disertai dengan kematian akibat *stroke* meningkat dua kali lipat setiap sepuluh tahun setelah umur 40 tahun. Satu dari tiga orang pada umur 80 tahun akan mengalami beberapa jenis *stroke* (Hutagaluh, 2019).

2. Jenis Kelamin

Riset menunjukkan bahwa Laki-laki lebih rentan terhadap *stroke* daripada perempuan dengan perbandingan 1,3:1. Namun dalam uraun batas menopause, perbandingan antara laki-laki dan wanita hampir sama. Perempuan yang meminum pil pengatur kehamilan dan menderita *stroke* juga lebih rentan. Oleh karena itu, lebih dari setengah korban *stroke* adalah perempuan (Hutagaluh, 2019).



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

3. Hipertensi

Hipertensi dapat menyebabkan kerusakan arteri di seluruh tubuh yang mengakibatkan pecahnya pembuluh darah dan sumbatan pada arteri di otak (Utama & Nainggolan, 2022).

4. Penyakit Jantung

Penyakit jantung berisiko besar dapat menyebabkan penderita *stroke* meninggal seperti *aritmia jantung*, *fibrilasi atrium*, *infark miokard*, dan gagal jantung diakibatkan karena *embolus* mengalir ke arteri yang mengarah ke otak dan bagian tubuh lainnya dan berdampak pada *infark* otak (Hutagalung, 2021).

5. Tingkat Glukosa

6. Obesitas

Pasien obesitas dengan indeks massa tubuh lebih dari 27,5 kg/m² memiliki risiko 1,5 kali lebih besar menderita *stroke* dan meninggal dunia dibandingkan pasien yang memiliki indeks massa tubuh normal yang tidak lebih dari 23,0 kg/m² (Hutagalung, 2021).

7. Merokok

Merokok merupakan salah satu faktor yang menyebabkan *stroke* dikarenakan perdarahan *intracerebral* yang merusak dinding arteri dan memecah pembuluh darah karena mengandung lebih dari 7.000 bahan kimia beracun (Utama & Nainggolan, 2022).

8. Penyalahgunaan obat / alkohol

Meminum alkohol lebih dari 2 gelas setiap hari meningkatkan kemungkinan terkena *stroke* sebesar 50%. Namun, hubungan antara konsumsi alkohol dan insiden *stroke* masih belum jelas. (Nurhalimah, Ratna Ningsih, 2024).

9. Kebiasaan Jenis Diet

Makanan dan minuman manis dapat menyebabkan peningkatan risiko *stroke* dikarenakan kebiasaan diet yang salah (Utama & Nainggolan, 2022).



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

10. Tingkat Kolesterol

Risiko penyakit *stroke* meningkat dikarenakan faktor riwayat penyakit jantung dan aterosklerosis. Nilai LDL adalah 100mg/dL jika nilai kolesterol ≥ 200 mg/dL, sedangkan nilai High Desity Lipoprotein (HDL) harus < 130 mg/dL(Kariasa, 2022).

2.5 **Confusion Matrix**

Menurut (Romadloni et al., 2022), *Confusion Matrix* digunakan untuk memberikan informasi tentang perbandingan hasil klasifikasi yang dilakukan oleh algoritma dengan hasil klasifikasi sebenarnya.

Tabel 1. *Confusion Matrix*

Confusion Matirx		Nilai Sebenarnya	
		<i>True</i>	<i>False</i>
Nilai Prediksi	<i>True</i>	TP	FP
	<i>False</i>	FN	TN

Hasil akurasi menggambarkan seberapa keakuratan model dalam mengklasifikasikan dengan benar.

$$\text{Akurasi} = \frac{(TP+TN)}{(TP+FP+FN+TN)} = \dots\dots\dots(5)$$

Presisi dapat menggambarkan keakuratan antara data yang diminta dengan hasil prediksi yang diberikan.

$$\text{Presisi} = \frac{(TP)}{(TP + FP)} = \dots\dots\dots(6)$$

Recall merupakan penggambaran dari keberhasilan model untuk menemukan kembali sebuah informasi.

$$\text{Recall} = \frac{TP}{(TP + FN)} = \dots\dots\dots(7)$$

Confusion matrix memberikan visualisasi akurasi *classifier* dengan membandingkan kelas aktual dan prediksi. Dalam dataset yang diuji, *biner classifier* memprediksi setiap *instance data* sebagai positif atau negatif. Terdapat empat hasil yaitu *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, *False Negative (FN)*. TP menghasilkan nilai yang diprediksi dengan benar sebagai positif *actual* dan FP menghasilkan nilai yang diprediksi salah menjadi positif



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

actual , misal, nilai-nilai negatif diprediksi sebagai positif, Sedangkan TN menghasilkan nilai yang diprediksi dengan benar sebagai negatif *actual* dan FN menghasilkan nilai positif namun diprediksikan negatif.(Sutoyo & Almaarif, 2020).

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 2.

Tabel 2. Penelitian Terkait

No	Peneliti	Judul	Kesimpulan
1	(Nuralia et al., 2023)	Implementasi <i>Naïve Bayes Classifier</i> Dalam Memprediksi Kelulusan Mahasiswa	Hasil penelitian menunjukkan algoritma <i>naive bayes</i> dapat digunakan untuk memprediksi kelulusan mahasiswa dengan menggunakan tiga <i>scenario</i> pengujian. Hasil akurasi tertinggi yaitu 95% dengan proporsi perbandingan 80:20 dengan nilai rata-rata <i>precision</i> , <i>recall</i> dan <i>fi-score</i> masing masing sebesar 95,16%, 95%, dan 95%.
2	(Veronica Agustin & Voutama, 2023)	Implementasi <i>Data Mining</i> Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan <i>Naïve Bayes</i>	Implementasi yang dilakukan menggunakan <i>dataset diabetes</i> perempuan diklasifikasi berdasarkan 8 atribut : <i>pregnancies</i> , <i>glucose</i> , <i>blood pressure</i> , <i>insulin</i> , <i>BMI</i> , <i>diabetes pedigree function</i> , <i>age</i> , dan <i>outcome</i> . Algoritma <i>naïve bayes</i> digunakan untuk mengklasifikasikan <i>dataset</i> perempuan yang menderita



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.

			<p><i>diabetes</i> dengan hasil nilai <i>Class precision</i> sebesar = 85.24%, dan nilai <i>Class recall</i> sebesar = 83.64%. Evaluasi pada klasifikasi menunjukkan nilai akurasi 78.50% yang artinya cukup baik tetapi nilai akurasi dapat ditingkatkan dengan cara atau <i>Assemble</i> yang lain. Kurva ROC (<i>Reciver Operating Characteristic</i>) yang dihasilkan berdasarkan klasifikasi permodelan <i>naïve bayes</i> dengan nilai AUC = 0.855 dan termasuk pada kategori <i>good</i> (baik).</p>
3	(Abidin et al., 2023)	<p>Klasifikasi Penyakit <i>Tuberkulosis</i> Menggunakan Metode <i>Naive Bayes</i> (Studi Kasus: Data Pasien Di Puskesmas Petung Kabupaten Penajam Paser Utara)</p>	<p>Menurut hasil dan pembahasan metode yang digunakan untuk mengklasifikasikan pasien <i>tuberkulosis</i> di Puskesmas Petung, Kabupaten Penajam Paser Utara, hasil akurasi untuk proporsi 90:10 sebesar 87,5% dan proporsi 70:30 sebesar 90,9%. Hasil evaluasi kedua proporsi menggunakan ROC <i>Curve</i> dapat dikategorikan sebagai <i>Excellent Classification</i>.</p>
4	(Paramitha et al., 2023)	<p>Klasifikasi Penyakit <i>Stroke</i> Menggunakan</p>	<p>Proses klasifikasi <i>Naive Bayes</i> untuk data pelatihan dan pengujian yang terdiri dari data pasien penyakit <i>stroke</i></p>



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

		Metode <i>Naïve Bayes</i>	dengan kelas terkena <i>stroke</i> atau tidak terkena <i>stroke</i> menggunakan 10 variabel bebas. Hasil evaluasi model menggunakan <i>confusion matrix</i> pada data penyakit <i>stroke</i> menunjukkan dengan proporsi 80:20 memiliki akurasi yang lebih tinggi dibandingkan dengan proporsi lainnya. Hasil evaluasi untuk proporsi 80:20 menunjukkan bahwa model tersebut terbaik. Penentuan pada proporsi pembagian antara data pelatihan dan data pengujian berdampak pada hasil presentase tingkat akurasi pengujian sebagai acuan untuk menentukan kelas pada data pengujian.
5	(Riany & Testiana, 2023b)	Penerapan Data Mining untuk Klasifikasi Penyakit <i>Stroke</i> Menggunakan Algoritma <i>Naïve Bayes</i>	Pengelompokan pasien menggunakan teknik klasifikasi <i>data mining</i> dan algoritma <i>naïve bayes</i> dilakukan untuk menentukan apakah pasien tersebut menderita <i>stroke</i> atau tidak. Berdasarkan pemrosesan pengujian dengan pembagian data 80 data pelatihan dan 20 data uji, terdapat 35 data terklasifikasi menderita penyakit <i>stroke</i> dan 962 data terklasifikasi tidak <i>stroke</i> dengan tingkat akurasi sebesar 92,48 dan menempatkan pada kategori <i>Good Classification</i> .



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

6	(Saragih et al., 2023)	Pembuatan Model Data Mining Untuk Memprediksi Infeksi Penyakit Covid-19 Dengan Menggunakan Algoritma <i>Naive Bayes Classifier</i>	Prediksi gejala kasus covid-19 menggunakan model <i>Naive Bayes Classifier</i> dan model <i>Gaussian Naive Bayes</i> menunjukkan kinerja yang sangat baik dengan nilai akurasi 95,73215599705665%, nilai presisi 0,99%, nilai recall 0,99%, dan nilai f-1 0,97%.
7	(Riany & Testiana, 2023a)	Penerapan Data Mining Untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma <i>Naive Bayes</i>	Penelitian menggunakan algoritma <i>naive bayes</i> bertujuan untuk menganalisis serta mengklasifikasi kumungkinan pasien penderita penyakit jantung koroner dalam sepuluh tahun mendatang . Terdapat 16 atribut penyebab penyakit jantung coroner diantaranya <i>gender, age, education,current smoker, cigs per day, BP Meds, prevalent stroke, prevalent hypertension, diabetes, total cholesterol, sys BP, dia BP, BMI, heart rate, dan glucose</i> . Tingkat akurasi yang didapatkan sebesar 79,10% dengan kategori cukup akurat..
8	(Rahma et al., 2023)	Implementasi Metode <i>Naive Bayes</i> untuk	Pengujian metode <i>Naive Bayes</i> yang menggunakan <i>10-Folds Cross Validation</i> untuk diagnosis penyakit


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.

		<p>Diagnosis Penyakit <i>Stroke</i></p>	<p><i>stroke</i> menghasilkan akurasi yang baik sebesar 82%. Terdapat 3 skenario penggunaan data awal, <i>undersampling</i>, dan <i>oversampling</i>. Hasil perbandingan menunjukkan tingkat akurasi, presisi, recall, dan <i>f1-score</i> masing-masing sebesar 82%, 82%, 97%, dan 88%. <i>Oversampling</i> dan <i>undersampling</i> digunakan karena data yang digunakan tidak seimbang sehingga mendapat tingkat akurasi, presisi, recall, dan <i>f1-score</i> masing-masing sebesar 64%, 65%, 60%, dan 61% dengan <i>oversampling</i>, hasilnya sebesar 71%, 57%, 52%, dan 52%. Dengan ini, penanganan data standar mendapatkan hasil lebih baik dibandingkan dengan penanganan data tidak seimbang.</p>
9	(Iis Dewi Ratih et al., 2022)	<p>Klasifikasi Penderita ISPA Menggunakan Metode <i>Naive Bayes Classifier</i></p>	<p>Penanganan untuk penderita ISPA menjadi tiga kategori yaitu ISPA ringan, ISPA sedang, dan ISPA berat. Hasil klasifikasi menggunakan metode <i>Naive Bayes Classifier</i> menunjukkan bahwa 18 orang adalah pasien ISPA kategori berat jumlah terbanyak dengan hasil akurasi pengujian <i>confussion matrix</i> sebesar 93.33% dan layak untuk dilakukan dalam klasifikasi ISPA.</p>



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.

10	(Zusrotun et al., 2022)	Sentimen Analisis Belajar Online Di Twitter Menggunakan Naïve Bayes	Keakuratan pada hasil pengujian algoritma <i>Naive Bayes Classifier</i> menunjukkan hasil akurasi sebesar 74,08%
11	(Bugis, 2022)	Metode <i>Naïve Bayes</i> Untuk Memprediksi Penyakit <i>Stroke</i>	Penelitian menggunakan teknik klasifikasi <i>naive bayes</i> merupakan teknik yang paling sederhana untuk memprediksi kemungkinan <i>stroke</i> . Hasil menunjukkan bahwa nilai akurasi sekitar 93.94% yang tergolong sangat tinggi menggunakan aplikasi <i>rapidminer</i> .
12	(Doni et al., 2021)	Penerapan <i>Data Mining</i> Untuk Klasifikasi Penyakit <i>Hepatocellular Carcinoma</i> Menggunakan Algoritma <i>Naïve Bayes</i>	Pencarian hasil yang optimal menggunakan algoritma <i>naive bayes</i> untuk mengklasifikasikan tingkat pasien penyakit <i>hepatocellular carcinoma</i> berdasarkan <i>Confusion Matrix</i> dan Kurva ROC menggunakan metode validasi <i>k – Fold Cross Validation</i> . Didapatkan hasil akurasi sebesar 70,30%, presisi sebesar 73,53%, dan recall sebesar 77,32% menggunakan metode validasi <i>10 – Fold Cross Validation</i> dengan nilai AUC (<i>Area Under Receiver Operating Characteristics Curve</i>) sebesar 0.783 dan termasuk dalam klasifikasi cukup.


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.

13	(Titimeidara & Hadikurniawati, 2021)	Implementasi Metode <i>Naive Bayes Classifier</i> Untuk Klasifikasi Status Gizi <i>Stunting</i> Pada Balita	Pengklasifikasian metode <i>Naive Bayes Classifier</i> untuk status gizi <i>stunting</i> pada balita berdasarkan jenis kelamin, umur, berat badan, tinggi badan, status miskin, dan kategori status gizi menunjukkan hasil akurasi sebesar 88% dari 300 data yang dibagi menjadi 275 data latih dan 25 data uji.
14	(Wijaya & Dwiasnati, 2020)	Implementasi <i>Data Mining</i> dengan Algoritma <i>Naive Bayes</i> pada Penjualan Obat	Penerapan algoritma <i>naive bayes</i> untuk klasifikasi penjualan obat disebuah apotek di Jakarta dengan <i>dataset</i> sebanyak 150 data, didapatkan hasil akurasi sebesar 88.00% menggunakan <i>tools Rapidminer</i> .
15	(Sutoyo & Almaarif, 2020)	Educational <i>Data Mining</i> untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme <i>Naive Bayes Classifier</i>	Penelitian dengan <i>dataset</i> yang diperoleh berasal dari Direktorat Sistem Informasi (SISFO) Universitas Telkom menggunakan <i>Naive Bayes Classifier</i> (NBC) , berisikan 4000 <i>instances</i> data dengan 6 atribut diantaranya Nilai Indeks Prestasi Semester (IPS) pada semester pertama, Nilai IPS semester kedua, Nilai IPS semester ketiga, Nilai IPS semester keempat, Nilai IPS semester keenam dan ditambah dengan class label kelulusan Tepat Waktu (TW) / Tidak Tepat Waktu (TTW) sebagai



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

			label output dengan hasil <i>accuracy</i> sebesar 73,725%, <i>precision</i> sebesar 0,742%, <i>recall</i> sebesar 0,736%, dan <i>F-measure</i> sebesar 0,735%.
16	(Afif, 2020)	Penerapan Algoritma <i>Naïve Bayes</i> Untuk Klasifikasi Penyakit <i>Diabetes Mellitus</i> di Rumah Sakit Aisyiah	Proses klasifikasi penyakit <i>diabetes mellitus</i> di Rumah Sakit Aisyiah menggunakan algoritma <i>naïve bayes</i> untuk mengklasifikasi 500 data <i>Early Stage Diabetes Risk Prediction Set</i> dengan 16 atribut yang berpengaruh, mendapatkan hasil <i>class precision</i> sebesar 82,35% dan <i>class recall</i> sebesar 87,50% serta tingkat akurasi 90,20% tergolong sebagai akurasi yang baik nilai AUC sebesar 0,955, yang menunjukkan <i>Excellent Classification</i> (Sangat Bagus).

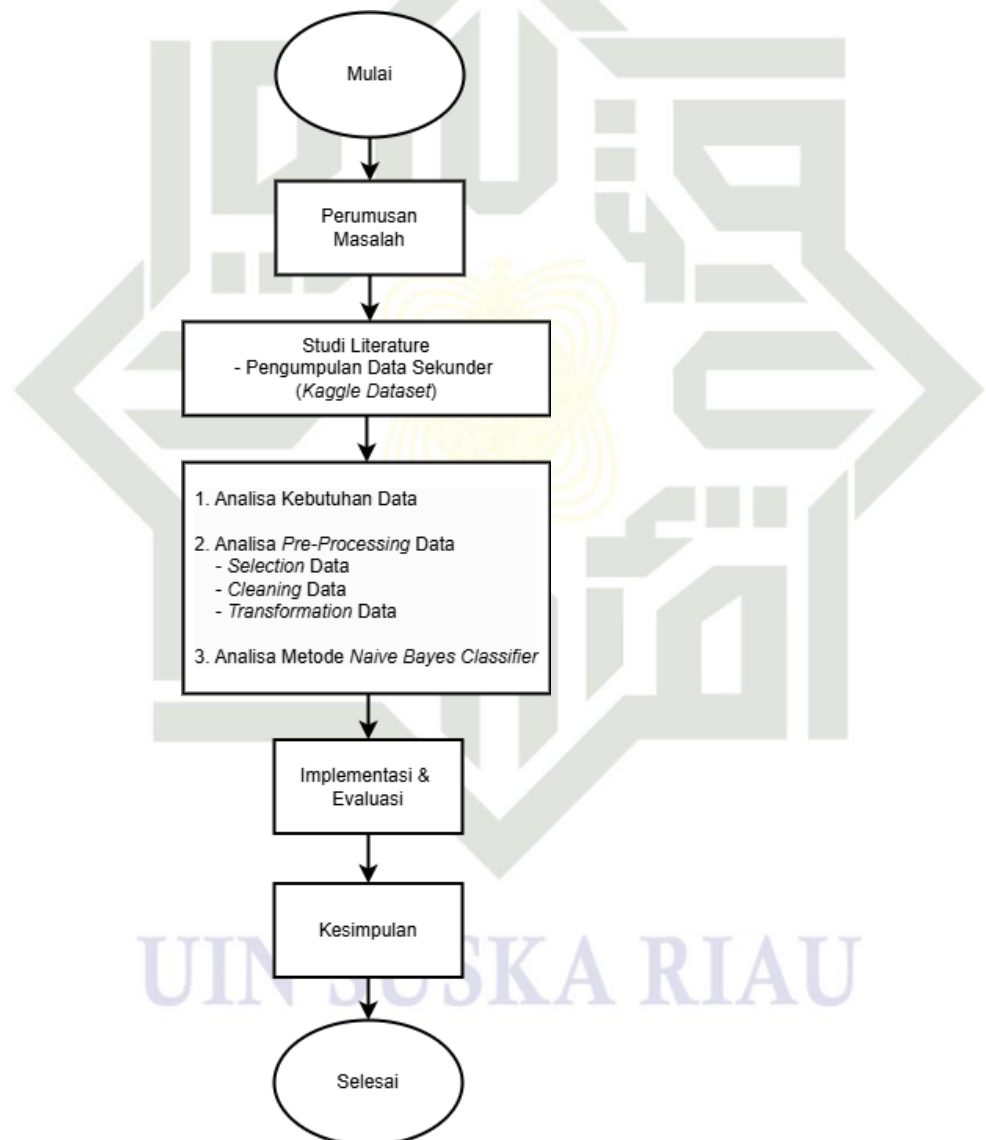
Hak Cipta Dilindungi Undang-Undang

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

BAB 3

METODOLOGI PENELITIAN

Metodologi penelitian merupakan tahapan yang sistematis secara ilmiah untuk mendapatkan tujuan dari skema yang jelas, terstruktur dan sistematis untuk memudahkan peneliti menyelesaikan penelitian ini. Adapun metodologi penelitian yang dilakukan dijabarkan sebagai berikut :



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

3.1 Perumusan Masalah

Pada perumusan masalah ini, peneliti menjelaskan tahapan awal dalam penelitian untuk memahami secara umum agar dapat mencari serta menemukan cara agar bisa menyelesaikan masalah tersebut. Permasalahan pada penelitian ini adalah bagaimana cara pemanfaatan *data mining* untuk mengklasifikasi data menggunakan algoritma *Naïve Bayes Classifier* untuk mengidentifikasi pasien penyakit *stroke*.

3.2 Pengumpulan Data

Data yang diperoleh untuk penelitian ini dengan cara pengumpulan data secara sekunder yang diambil dari *Kaggle*. Data dapat diakses melalui <https://www.kaggle.com/datasets/teamincirbo/stroke-prediction/data>. Adapun bentuk dataset yang digunakan dapat dilihat pada Tabel 3.

Tabel 3. Dataset Stroke

Patient ID	Patient Name	Age	Gender	Hyper tension	Heart Disease	Marital Status	Work Type	Residence Type	...	Diagnosis
18153	Mamooty Khurana	56	Male	0	1	Married	Self-employed	Rural	...	Stroke
62749	Kaira Subramaniam	80	Male	0	0	Single	Self-employed	Urban	...	Stroke
32145	Dhanush Balan	26	Male	1	1	Married	Never Worked	Rural	...	Stroke
6154	Ivana Baral	73	Male	0	0	Married	Never Worked	Urban	...	No Stroke
48973	Darshit Jayaraman	51	Male	1	1	Divorced	Self-employed	Urban	...	Stroke
29307	Advika Kota	62	Female	0	0	Single	Private	Urban	...	Stroke
25525	Elakshi Karan	40	Female	1	0	Married	Private	Urban	...	No Stroke
...
7372	Raghav Handa	72	Female	1	1	Married	Self-employed	Rural	...	Stroke

Gambar di atas menunjukkan *dataset* awal pada penyakit *stroke*, *dataset* penyakit *stroke* ini berjumlah 15000 data pasien *stroke* yang akan dilabel menjadi kelas yaitu *stroke* dan *no stroke*.

3.3 Analisa Data

Tujuan penelitian pada tahap analisa data yaitu untuk menguraikan detail mengenai gambaran data yang digunakan pada 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.

3.3.1 Analisa Kebutuhan Data

Tahapan awal untuk menganalisa *dataset* pasien penyakit stoke yaitu menganalisa seluruh kebutuhan data yang di peroleh dengan cara pengumpulan data sekunder dari *Kaggle*.

3.3.1.1 Selection Data

Pemilihan atribut yang digunakan pada *dataset* penyakit *stroke* yang berasal dari *kaggle dataset* terdiri dari 15000 data dengan 13 atribut fitur dan 1 label kelas yang sudah diseleksi oleh dokter spesialis saraf RS Awal Bross Pekanbaru yaitu *Age, Gender, Hypertension, Heart Disease, Average Glucose Level, Body Mass Index (BMI), Smoking Status, Alcohol Intake, Dietary Habits, HDL, LDL, Systolic, Diastolic, Diagnosis*.

3.3.1.2 Cleaning Data

Cleaning data adalah tahapan pertama yang akan dilakukan pada data yang tersedia dengan proses menyeleksi, memperbaiki maupun menghapus data. Pembersihan data dilakukan dalam data *stroke* apabila data digunakan tidak konsisten, hilang atau nilai *NaN* dan *Unknown*. Adapun beberapa langkah yang dilakukan untuk mengatasi masalah tersebut :

1. Missing Value

Missing value merupakan data yang tidak diketahui atau hilang yang tidak berada pada dataset. Penanganan yang dilakukan pada penelitian dengan 2 cara yaitu : menghapus baris / kolom yang hilang atau mengganti nilai yang hilang.

2. Duplicate Data

Duplicate data adalah data yang kemunculannya itu lebih dari sekali pada dataset. Penanganan yang dilakukan pada penelitian yaitu menghapus data duplicate tersebut dari dataset.



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

3. Inconsistent Data

Inconsistent data (data yang tidak konsisten) yaitu data pada dataset yang memiliki format yang berbeda untuk nilai yang sama atau kesalahan penulisan. Penanganan yang dilakukan pada penelitian yaitu melakukan normalisasi data, koreksi kesalahan penulisan, dan validasi data.

4. Outlier

Outlier yaitu data yang nilainya jauh berbeda dari nilai lainnya pada dataset. Penanganan yang dilakukan pada penelitian yaitu menghapus outlier, melakukan tranformasi data, dan menerapkan winsorizing.

3.3.1.3 Transformasi Data

Transformasi data merupakan tahapan merubah format data dalam bentuk yang sesuai untuk proses data mining. Pada tahap ini, peneliti akan merubah format data menjadi format data kategorikal agar bisa digunakan pada metode *naïve bayes* yang ditampilkan pada Tabel 4.

Tabel 4. Transformasi Data

No	Atribut	Transformasi Atribut
1	Age	Menjadi 3 kategori yaitu Young (18 - 44 tahun), Adult (45 - 64 tahun), Elderly (65 - 90 tahun)
2	Hypertension	Menjadi 2 kategori yaitu No Hypertension (0), Hypertension (1)
3	Heart Disease	Menjadi 2 kategori yaitu No Heart Disease (0), Heart Disease (1)
4	Average Glucose Level	Menjadi 3 kategori yaitu Normal (≤ 130.91), Diabetes ($130.91 - 183.73$), Pradiabetes
5	Body Mass Index (BMI)	Menjadi 4 kategori yaitu Underweight (< 18.5), Normal ($18.5 - 24$), Overweight ($25 - 29$), Obesitas
6	Systolic	Menjadi 5 kategori yaitu Normal (< 120), Elevated ($120 - 129$), Hypertension Stage 1 ($130 - 139$), Hypertension Stage 2 ($140 - 179$), Hypertension Stage 3 (≥ 180)

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

Diastolic	Menjadi 5 kategori yaitu Normal (< 80), Elevated (80 - 89), Hypertension stage 1 (90 - 99), Hypertension Stage 2 (100 - 119), Hypertension Stage 3 (\geq 120)
HDL	Menjadi 3 kategori yaitu Low (< 40), Normal (40 - 60), High
LDL	Menjadi 4 kategori yaitu Optimal (< 100), Near Optimal (100 - 130), Borderline High (131 - 160), High (161 - 190), Very High

Pada tabel diatas terdapat 4 atribut baru yang mana atribut ini didapat dari transformasi 2 atribut :

1. Atribut *Blood Pressure Levels*

Hasil wawancara bersama dokter spesialis saraf di RS Awal Bross Pekanbaru, *Blood Pressure Levels* ini dibagi menjadi 2 atribut yaitu *Systolic* (tekanan darah saat jantung memompa darah) dan *Diastolic* (tekanan darah saat jantung beristirahat). Pembagian ini membantu memahami kondisi tekanan darah pasien dengan lebih jelas.

2. Atribut *Cholesterol Levels*

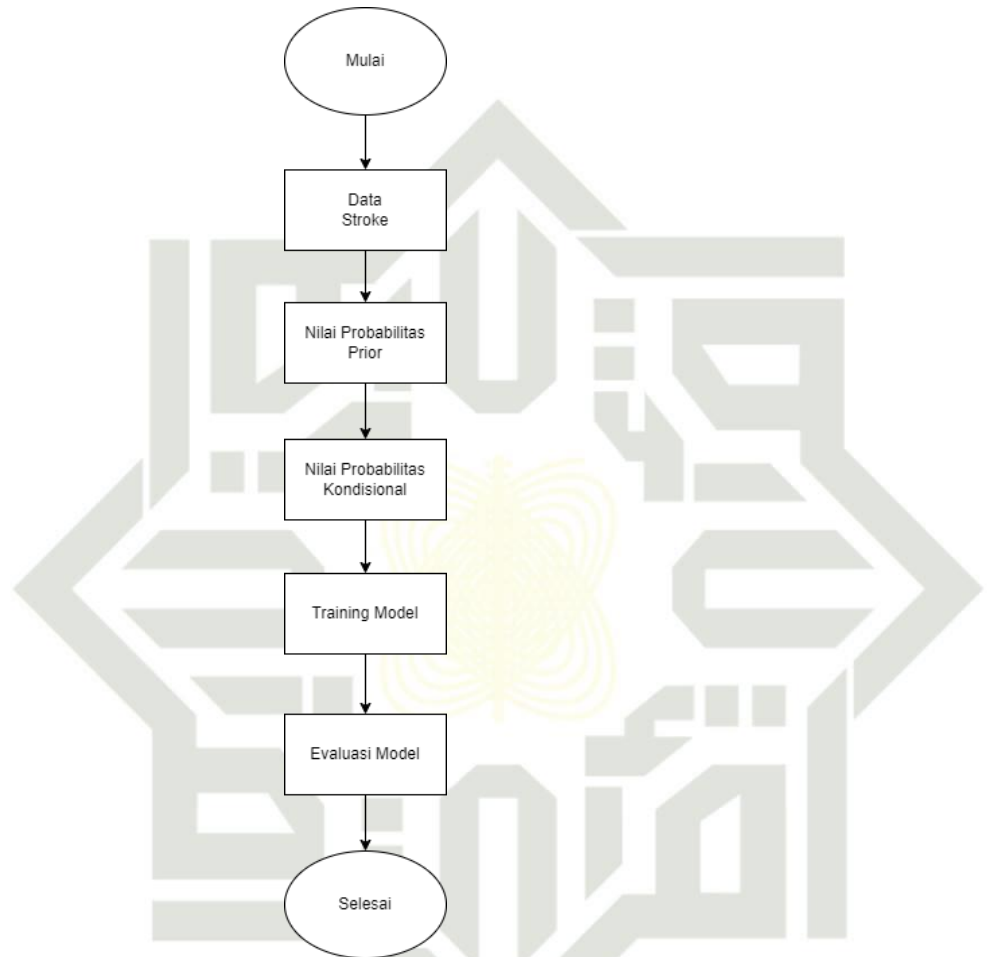
Hasil wawancara bersama dokter spesialis saraf di RS Awal Bross Pekanbaru, *Cholesterol Levels* ini dibagi menjadi HDL yaitu kolesterol baik dan LDL yaitu kolesterol jahat untuk memahami risiko kesehatan dengan lebih baik.

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.

3.2 Naïve Bayes Classifier

Pada tahap ini, peneliti akan menerapkan algoritma *Naïve Bayes Classifier* dengan bentuk *flowchart* atau tahapan yang diterapkan seperti berikut:



Gambar 3. Flowchart Naïve Bayes Classifier

Berikut ini adalah tahapan dari proses *Naïve Bayes Classifier* berdasarkan *flowchart* diatas :

1. Nilai Probabilitas Prior

Setelah dilakukannya tahapan *preprocessing data* dan pembagian dataset kedalam data uji dan data latih, maka dilakukanlah perhitungan probabilitas awal dari setiap kelas. Seperti contoh $P(\text{Stroke})$ dan $P(\text{No Stroke})$. Untuk menghitung nilai probabilitas prior, dapat dirujuk pada Rumus (2).

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. Nilai Probabilitas Kondisional

Tahapan selanjutnya yaitu menghitung nilai probabilitas dari setiap fitur yang telah diberikan kelas tertentu. Seperti, menghitung $P(\text{Umur}|\text{Stroke})$, $P(\text{Jenis Kelamin}|\text{Stroke})$. Untuk menghitung nilai probabilitas kondisional, dapat dirujuk pada Rumus (3).

3. Training Model

Pada tahap ini, model Naive Bayes dilatih menggunakan data pelatihan. Probabilitas prior dan kondisional yang telah dihitung disimpan sebagai bagian dari model. Algoritma Naive Bayes mengasumsikan bahwa setiap fitur bersifat independen, sehingga probabilitas keseluruhan adalah produk dari probabilitas individu fitur. Untuk perhitungan pengujian model data pelatihan, dapat dirujuk pada Rumus (4).

4. Evaluasi Model

Model yang telah dilatih kemudian diuji menggunakan data pengujian yang belum pernah dilihat selama proses pelatihan. Dalam tahap evaluasi ini, berbagai metrik dihitung, termasuk akurasi, presisi dan recall. Metrik-metrik ini digunakan untuk menilai kinerja model, membantu kita memahami seberapa baik model dapat mengklasifikasikan data baru dengan tepat. Untuk perhitungan mengenai akurasi, presisi dan recall pada tahapan evaluasi model, dapat dirujuk pada Rumus (5), (6), dan (7).

4.4 Implementasi dan Pengujian

Tahapan implementasi dan pengujian pada penelitian ini diperlukan sebagai prosedur secara sistematis agar desain pada dokumen dapat terselesaikan dan mengevaluasi kinerja model agar analisis lebih menyeluruh sehingga mendapatkan tingkat akurasi.



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

3.4.1 Implementasi

Implementasi dibutuhkan sebagai prosedur tersistematika yang dilakukan dalam menyelesaikan desain pada dokumen yang disetujui. Penelitian ini membutuhkan perangkat pendukung seperti perangkat keras dan perangkat lunak sebagai berikut :

Perangkat Keras (Hardware)

- a. CPU : Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz ~2.0GHz
- b. Memori (RAM) : 12 GB
- c. SSD : 256 GB

Perangkat Lunak (Software)

- a. Platform : Windows 10 Home Single Language 64-bit
- b. Tools : Google Collab

3.4.2 Pengujian

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

✓ Pengujian Akurasi

Pengujian menggunakan *Confussion matrix* salah satu cara yang sering digunakan untuk pemrosesan evaluasi model dengan memprediksi kebenaran objek serta dapat membantu dalam mengevaluasi kinerja model dan membuat analisis yang lebih menyeluruh dan tidak hanya menghitung probabilitas klasifikasi yang sebenarnya (akurasi) yang dapat dirujuk pada persamaan (5).

3.5 Kesimpulan dan Saran

Berdasarkan pengujian yang telah dilakukan, peneliti menarik kesimpulan yang mengacu untuk rumusan masalah dan tujuan penelitian. Saran untuk penelitian yang sedang dilakukan yaitu dapat mengembangkan penelitian selanjutnya dengan peningkatan kualitas penelitian.



BAB 5

PENUTUP

5.1 Kesimpulan

Dari hasil implementasi dan pengujian yang telah dilakukan, maka peneliti mendapat kesimpulan sebagai berikut :

Penerapan metode *Naïve Bayes Classifier* untuk mengklasifikasi penyakit *stroke* berdasarkan data pasien *stroke kaggle* berhasil diterapkan dengan baik. Hasil Akurasi tertinggi dari perbandingan evaluasi beberapa pengujian pada klasifikasi penyakit *stroke* menggunakan metode *Naïve Bayes Classifier* dari dataset filter dokter spesialis syaraf RS Awal Bros Pekanbaru, Keseluruhan Atribut Data, *Feature Selection*, *K-Fold*, *RapidMiner*, didapatkan hasil akurasi tertinggi menggunakan *Chi-Square Feature Selection* dengan jumlah data sebanyak 12500 data dengan 9 atribut sebesar 53 %, presisi 53%, dan nilai recall 53%.

5.2 Saran

Dari kesimpulan yang didapatkan pada penelitian ini, diharapkan untuk penelitian selanjutnya sebagai berikut :

Diharapkan untuk penelitian selanjutnya, dapat menggunakan beberapa optimasi yang relevan seperti *cross-validation* atau *handling imbalanced data* menggunakan *SMOTE* agar pengklasifikasian penyakit *stroke* mendapatkan tingkat akurasi yang lebih baik dan berguna untuk memprediksi penyakit *stroke*.

Penelitian selanjutnya disarankan untuk melihat visualisasi dan persebaran data pasien *stroke* serta lakukan proses clustering seperti *K-Means* atau *DBSCAN* dalam pengelompokkan data untuk fitur tambahan model prediktif.

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

DAFTAR PUSTAKA

- Abidin, A. A., Goejantoro, R., & Fathurahman, M. (2023). Klasifikasi Penyakit Tuberkulosis Menggunakan Metode Naive Bayes (Studi Kasus: Data Pasien Di Puskesmas Petung Kabupaten Penajam Paser Utara). *Ekspensial*, 14(1), 11. <https://doi.org/10.30872/ekspensial.v14i1.1031>
- Afif, A. (2020). Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus di Rumah Sakit Aisyiah. *Jurnal Ilmu Komputer Dan Matematika*, 1(1), 40–46.
- 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>
- Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>
- Bugis, H. (2022). Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 6(1), 8–14. <https://doi.org/10.47970/siskom-kb.v6i1.317>
- Doni, B. T. R., Susanti, S., & Mubarak, A. (2021). Penerapan Data Mining Untuk Klasifikasi Penyakit Hepatocellular Carcinoma Menggunakan Algoritma Naïve Bayes. *Jurnal Responsif : Riset Sains Dan Informatika*, 3(1), 12–19. <https://doi.org/10.51977/jti.v3i1.403>
- Dwilaksono, D., Fau, T. E., Siahaan, S. E., Siahaan, C. S. P. B., Karo, K. S. P. B., & Nababan, T. (2023). Faktor-Faktor yang Berhubungan dengan Terjadinya Stroke Iskemik pada Penderita Rawat Inap. *Jurnal Penelitian Perawat Profesional*, 5(2), 449–458. <https://doi.org/10.37287/jppp.v5i2.1433>

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

- Padri, W. (2023). Klasifikasi Penyakit Hati dengan Menggunakan Metode Naive Bayes. *Jurnal Informasi Dan Teknologi*, 5(1), 32–36. <https://doi.org/10.37034/jidt.v5i1.230>
- Gustin Rahayu, T. (2023). Analisis Faktor Risiko Terjadinya Stroke Serta Tipe Stroke. *Faletahan Health Journal*, 10(1), 48–95. www.journal.lppm-stikesfa.ac.id/ojs/index.php/FHJ
- Haiga, Y., Prima Putri Salman, I., & Wahyuni, S. (2022). Perbedaan Diagnosis Stroke Iskemik dan Stroke Hemoragik dengan Hasil Transcranial Doppler di RSUP Dr. M. Djamil Padang. *Scientific Journal*, 1(5), 391–400. <https://doi.org/10.56260/sciena.v1i5.72>
- Hutagaluh, M. S. (2019). *Panduan Lengkap Stroke: Mencegah, Mengobati dan Menyembuhkan*. Nusamedia. <https://books.google.co.id/books?id=UmVcEAAAQBAJ>
- Hutagalung, M. S. (2021). *Penyebab Kematian pada Pasien Stroke serta Peran Keluarga dalam Membantu Proses Penyembuhan Stroke: Panduan Lengkap Stroke*. Nusamedia. <https://books.google.co.id/books?id=uKNsEAAAQBAJ>
- Ratih, S M Retnaningsih, & V M Dewi. (2022). Klasifikasi Kualitas Tanah Menggunakan Metode Naive Bayes Classifier. *Jurnal Aplikasi Matematika Dan Statistik*, 1(1), 11–20. <https://doi.org/10.53625/jams.v1i1.4427>
- Irmayani, W. (2021). Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi. *Jurnal Khatulistiwa Informatika*, IX(1), 68–72.
- Indah Pratiwi, Dwiarti Rahma Utami, & Enri, U. (2022). Klasifikasi Data Mining Dalam Menentukan Pemberian Pinjaman Berbasis Arisan Online dengan Algoritma C4.5. *InfoTekJar : Jurnal Nasional Informatika Dan Teknologi Jaringan*, 6(2), 261–266. <https://doi.org/10.30743/infotekjar.v6i2.4745>
- Journal, A., Rizki, F., Herdian Andika, T., Alfarisi Salimu, S., Murfid Kusuma, I., Informatika, T., Teknologi Informasi, P., Informasi, S., Perangkat Lunak, R.,

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

Teknologi dan Informatika, F., Keguruan dan Ilmu Pendidikan, F., Aisyah Pringsewu, U., & Pringsewu, S. (2022). Penggunaan Data Mining Sebagai Pengambilan Keputusan Penerimaan Bantuan Terhadap Rumah Ibadah (Studi Kasus : Rumah Ibadah Di Kabupaten Tanggamus). *Aisyah Journal Of Informatics and Electrical Engineering*, 4(1), 55–65. <https://jti.aisyahuniversity.ac.id/index.php/AJIEE/article/view/105>

Kariasa, I. M. (2022). *Antisipasi Serangan Stroke Berulang*. Penerbit NEM. <https://books.google.co.id/books?id=jdiAEAAAQBAJ>

Nuralia, S., Harliana, H., & Prabowo, T. (2023). Implementasi Naïve Bayes Classifier Dalam Memprediksi Kelulusan Mahasiswa. *Journal Automation Computer Information System*, 3(1), 63–72. <https://doi.org/10.47134/jacis.v3i1.57>

Nurhalimah, Ratna Ningsih, Y. M. (2024). *TERAPI KOGNITIF PADA PASIEN STROKE*. Penerbit P4I. <https://books.google.co.id/books?id=au4CEQAAQBAJ>

Paramitha, Y. N., Nuryaman, A., Faisol, A., Setiawan, E., & Nurvazly, D. E. (2023). Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes. *Jurnal Siger Matematika*, 04(01), 11–16. <https://www.kaggle.com/datasets/zzettrkalkpakbal/full-filled->

Ruspitawuri, A., Santoso, E., & Dewi, C. (2019). Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode K-Nearest Neighbor dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(4), 3319–3324. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4916>

Putra Kusuma, A., Tri Utami, I., & Purwono, J. (2022). Pengaruh Terapi “Menggengam Bola Karet Bergerigi” Terhadap Perubahan Kekuatan Otot Pada Pasien Stroke Diukur Menggunakan Hangryp Dynamometer Di Ruang Syaraf Rsud Jend a Yani Kota Metro. *Jurnal Cendikia Muda*, 2(1), 17–23. <https://journal.uny.ac.id/index.php/jpts/article/view/53930>



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

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

- Rahayu, P. W., Sudipa, I. G. I., Suryani, S., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M. N., Slamet, I., Harlina, S., & Maysanjaya, I. M. D. (2024). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia. <https://books.google.co.id/books?id=vCruEAAAQBAJ>
- Rahma, A., Nisa', S., Nugroho, H., & Yuliasuti, G. E. (2023). *Implementasi Metode Naïve Bayes untuk Diagnosis Penyakit Stroke*. 177–184. <https://ejournal.itats.ac.id/snestikdanhttps://snestik.itats.ac.id>
- Riany, A. F., & Testiana, G. (2023a). Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes. *MDP Student Conference*, 2(1), 297–305. <https://doi.org/10.35957/mdp-sc.v2i1.4388>
- Riany, A. F., & Testiana, G. (2023b). Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes. *Jurnal SAINTEKOM*, 13(1), 42–54. <https://doi.org/10.33020/saintekom.v13i1.352>
- Romadloni, P., Adhi Kusuma, B., & Maulana Baihaqi, W. (2022). Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 622–628. <https://doi.org/10.36040/jati.v6i2.5238>
- Saragih, Y. W., Lumbantoruan, S. K., & ... (2023). Pembuatan Model Data Mining Untuk Memprediksi Infeksi Penyakit Covid 19 Dengan Menggunakan Algoritma Naive Bayes Classifier. *Jurnal ...*, 3(Juli), 187–193. <http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta/article/view/1114>
<http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta/article/download/1114/733>
- Sari, R. M., & M, Y. A. (2024). *Klasifikasi Data Mining*. Serasi Media Teknologi. <https://books.google.co.id/books?id=xTwXEQAAQBAJ>
- Sianturi, F. A., Hasugian, P. M., Simangunsong, A., Nadeak, B., & Sihotang, H. T. (2019). *DATA MINING: Teori dan Aplikasi Weka*. IOCS Publisher.


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

<https://books.google.co.id/books?id=MWcHEAAAQBAJ>

Sulianta, F. (2023). *Basic Data Mining from A to Z*. Feri Sulianta. <https://books.google.co.id/books?id=JcLhEAAAQBAJ>

Ututoyo, E., & Almaarif, A. (2020). Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1), 95–101. <https://doi.org/10.29207/RESTI.V4I1.1502>

Pitimeidara, M. Y., & Hadikurniawati, W. (2021). Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita. *Jurnal Ilmiah Informatika*, 9(01), 54–59. <https://doi.org/10.33884/jif.v9i01.3741>

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>

Veronica Agustin, A., & Voutama, A. (2023). Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1002–1007. <https://doi.org/10.36040/jati.v7i2.6808>

Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1–7. <https://doi.org/10.31311/ji.v7i1.6203>

Muli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.

Susrotun, O. P., Murti, A. C., & Fiati, R. (2022). Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(3), 310–319. <https://doi.org/10.23887/janapati.v11i3.49160>



DAFTAR RIWAYAT HIDUP



Informasi Personal

Nama : RAYHAN ATHALLAH IRSAN
 Tempat/Tanggal Lahir : Pekanbaru, 11 Desember 2001
 Alamat : Jl. Swadaya II No.23, Tj. Rhu, Kec. Lima Puluh, Kota Pekanbaru
 Jenis Kelamin : Laki - Laki
 Agama : Islam
 Tinggi badan : 175 cm
 Kewarganegaraan : Indonesia
 Hobby : Olahraga, Nonton & Main Game
 Motto : “Tetap rendah hati dan jangan menyerah”

Riwayat Pendidikan

Tahun 2008-2014 : SDN 17 Pekanbaru
 Tahun 2014-2017 : MTsN 3 Pekanbaru
 Tahun 2017-2020 : SMKN 2 Pekanbaru

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