PENERAPAN DATA MINING DENGAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PENYAKIT STROKE OTAK

PROPOSAL TUGAS AKHIR

Disusun Oleh :

Zurman Rivliansyah

8020190045

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir

Sebagai Akhir Proses Studi Strata 1

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS DINAMIKA BANGSA

2022

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : Penerapan Data Mining Dengan Algoritma C4.5 Untuk Memprediksi Penyakit Stroke Otak

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

a. Nama Lengkap : Zurman Rivliansyah
b. NIM : 8020190045
c. Jenis Kelamin : Laki-Laki
d. Tempat/Tgl Lahir : Jambi/05 Agustus 2001
e. Alamat : Perumahan Anugrah Argenta 4, Block A, No.8, RT.06, Lrg. Kelapa Sawit, Jerambah Bolong, Kab. Muaro Jambi
f. No. Telepon : 081310619713

Email : zurmanhopkins01@gmail.com

ABSTRAK

Stroke tergolong jenis penyakit yang paling banyak diderita masyarakat indonesia dan memiliki tingkat kematian yang tinggi, stroke berperan merusak ataupun mengganggu fungsi otak yang dimana aliran darah yang mengalir untuk otak manusia terganggu dan hal ini memungkinkan pasien yang terkena penyakit stroke otak akan mengalami kelumpuhan atau kematian. Stroke pada otak tidak bisa dianggap penyakit biasa karena jika dilihat dari gejala dan akibat dari stroke otak, penyakit ini harus dianggap sebagai penyakit serius. Informasi tentang gejala ataupun penyebab penyakit stroke otak sangatlah penting untuk diketahui untuk dapat mengambil langkah pencegahan terhadap penyakit tersebut, Hal ini bisa dilakukan dengan dengan menganalisa data penyakit stroke dengan teknik data mining yaitu klasifikasi, Klasifikasi berperan untuk menemukan suatu bentuk model dan membedakan suatu kelas dari konsep data, setelah itu model digunakan untuk memprediksi label dari kelas objek yang status label kelasnya belum diketahui.

Kata Kunci : Penerapan, Algoritma, *C4.5*, Prediksi, Stroke, Otak

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul “Penerapan Data Mining Dengan Algoritma C4.5 Untuk Memprediksi Penyakit Stroke Otak” dengan sebaik-baiknya.

Penulis menyadari bahwa dalam penulisan proposal skripsi ini banyak mengalami kendala dan kekurangan, namun berkat bantuan, bimbingan, dan petunjuk dari banyak pihak sehingga kendala-kendala yang dihadapi tersebut dapat diatasi dan penulis dapat menyelesaikan proposal skripsi ini dengan baik.

Penulis berharap agar proposal skripsi ini dapat bermanfaat bagi kita semua. Akhir kata penulis mengucapkan terima kasih.

Jambi, September 2022

Penulis

DAFTAR ISI

IDENTITAS PROPOSAL PENELITIAN ..... **ii**

ABSTRAK ..... **iii**

KATA PENGANTAR ..... **iv**

DAFTAR ISI ..... **v**

BAB I PENDAHULUAN **1**

1.1. LATAR BELAKANG 1

1.2. RUMUSAN MASALAH 2

1.3. BATASAN MASALAH 2

1.4. TUJUAN DAN MANFAAT PENELITIAN 2

1.4.1. Tujuan Penelitian 2

1.4.2. Manfaat Penelitian 2

BAB II LANDASAN TEORI **4**

2.1. DEFINISI PENYAKIT 4

2.2. DATA MINING 4

2.2.1. Tahapan Proses dalam Data Mining 5

2.2.2. Pengelompokkan Data Mining 8

2.3. KLASIFIKASI 10

2.4. DECISION TREE & ALGORITMA C4.5 10

2.5. PENELITIAN SEJENIS 11

BAB III METODOLOGI PENELITIAN **16**

3.1. KERANGKA KERJA PENELITIAN 16

3.2. ANALISA TEKNIK ALGORITMA C4.5 18

3.3. METODE PENELITIAN 18

3.4. ALAT PENELITIAN 19

3.5. JADWAL PENELITIAN 19

DAFTAR PUSTAKA

LAMPIRAN DATASET

BAB I
PENDAHULUAN

1.1. LATAR BELAKANG

Kesehatan merupakan sesuatu yang penting untuk diperhatikan, seseorang yang meremehkan kesehatan akan mudah terkena berbagai macam jenis penyakit dan cara terbaik untuk menghindari hal tersebut adalah dengan cara mencegahnya. Penyakit memiliki banyak jenis dan akibat yang ditimbulkan dari penyakit tersebut ada bermacam-macam, terdapat penyakit yang memungkinkan penderitanya mengalami kematian salah satunya yaitu penyakit yang dapat menyebabkan kematian adalah stroke.

Stroke tergolong jenis penyakit yang paling banyak diderita masyarakat indonesia dan memiliki tingkat kematian yang tinggi [1], stroke berperan merusak ataupun mengganggu fungsi otak yang dimana aliran darah yang mengalir untuk otak manusia terganggu dan hal ini memungkinkan pasien yang terkena penyakit stroke otak akan mengalami kelumpuhan atau kematian. Stroke otak tidak hanya dialami oleh orang dewasa ataupun yang sudah tua, akan tetapi stroke otak juga bisa dialami oleh remaja ataupun anak-anak sekalipun. Stroke pada otak tidak bisa dianggap penyakit biasa karena jika dilihat dari gejala dan akibat dari stroke otak, penyakit ini harus dianggap sebagai penyakit serius.

Informasi tentang gejala ataupun penyebab penyakit stroke otak sangatlah penting untuk diketahui untuk dapat mengambil langkah pencegahan terhadap penyakit tersebut.

Sehubungan dengan permasalahan diatas, penulis tertarik untuk melakukan penelitian yang berjudul **"*Penerapan Data Mining Dengan Algoritma C4.5 Untuk Memprediksi Penyakit Stroke Otak*"**

1.2. RUMUSAN MASALAH

Berdasarkan dari latar belakang, maka didapat rumusan masalah sebagai berikut :

1. Bagaimana penerapan *Algoritma C4.5* dalam klasifikasi penyakit Stroke Otak?
2. Seberapa besar tingkat akurasi *Algoritma C4.5* jika digunakan untuk mengklasifikasikan penyakit Stroke Otak?

1.3. BATASAN MASALAH

Berdasarkan dari latar belakang, maka adapun batasan masalah dalam penelitian ini, yaitu:

1. Data yang diperlukan untuk pengujian data adalah dataset penderita penyakit Stroke Otak.
2. Variabel yang digunakan dalam penelitian ini terdapat *Gender, Age, Hypertension, Heart Disease, Ever Maried, Work Type, Residence Type, AVG Glucose Level, BMI, Smoking Status.*

1.4. TUJUAN DAN MANFAAT PENELITIAN

1.4.1. Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu :

1. Mengklasifikasi data penderita penyakit Stroke Otak.
2. Mendapatkan akurasi yang tepat dan maksimal untuk melakukan klasifikasi penyakit Stroke Otak dengan menerapkan *Algoritma C4.5*.
3. Mengetahui sebeberapa besar tingkat akurasi *Algoritma C4.5* dalam klasifikasi penyakit Stroke Otak.

1.4.2. Manfaat Penelitian

Adapun beberapa manfaat yang akan didapat dalam melakukan penelitian ini, yaitu :

1. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi *data mining* khususnya untuk klasifikasi dataset dunia kesehatan.

2. Memberikan informasi mengenai gejala dan penyakit Stroke agar mampu mengetahui penyebab dan akibat dari penyakit Stroke sehingga mampu mengambil langkah pencegahan lebih awal.
3. Menambah wawasan bagi penulis dalam pengklasifikasian dataset dunia kesehatan.

BAB II
LANDASAN TEORI

2.1. DEFINISI PENYAKIT

Penyakit merupakan sesuatu yang membuat fungsi pada tubuh tidak berjalan dengan baik atau tidak normal, ada terdapat berbagai macam jenis penyakit yang dapat mengganggu fungsi organ, struktur, dan sistem pada tubuh, penyakit juga terdapat yang menular dan tidak menular [2]

2.2. DATA MINING

Data mining adalah kegiatan menambang suatu data dengan cara mengolah data dengan metode tertentu dan tentunya juga dengan menerapkan ilmu pada bidang lain seperti Matematika dan Statistik, penerapan teknik data mining sangat baik untuk melakukan analisis pada data yang berjumlah besar untuk mencari atau menentukan pola penting [3]

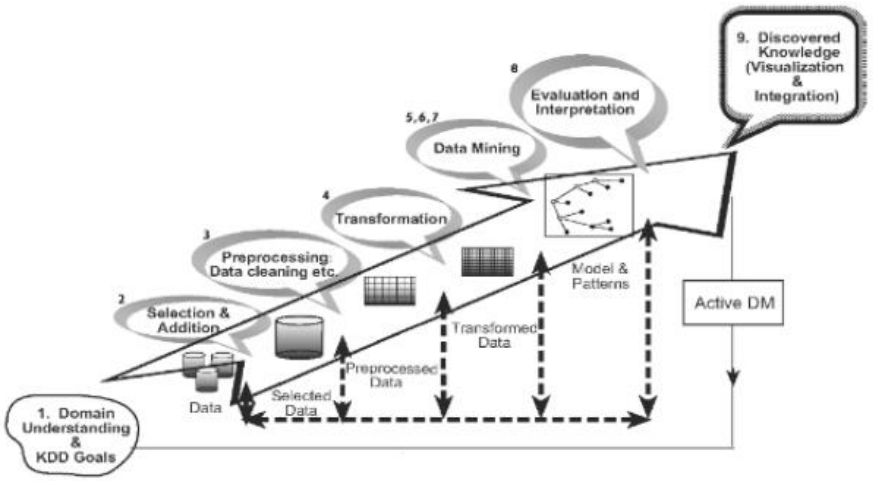
Menurut Turban et al. [4], "*data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatik yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar".

Berdasarkan definisi – definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah :

1. *Data Mining* adalah kegiatan menambang suatu data dengan cara mengolah data dengan metode tertentu.
2. Penerapan teknik data mining sangat baik untuk melakukan analisis pada data yang berjumlah besar.

2.2.1. Tahapan Proses dalam *Data Mining*

Ada beberapa tahapan dalam proses *data mining*. Diagram dibawah ini menggambarkan beberapa tahap/proses yang berlangsung dalam *data mining*.



Gambar 2.1. Fase – Fase dalam *Data Mining* [5]

1. **Pembentukan pemahaman domain aplikasi**. Pada tahap ini menentukan tujuan dari end-user dan bagian terkait dimana KDD dilakukan. Mengembangkan pemahaman tentang domain aplikasi ini adalah awal langkah persiapan. Ini mempersiapkan adegan untuk memahami apa yang harus dilakukan dengan banyak keputusan (tentang transformasi, algoritma, representasi, dll). Orang-orang yang bertanggung jawab atas proyek KDD perlu memahami dan menentukan tujuan dari pengguna akhir dan lingkungan di mana pengetahuan penemuan proses akan berlangsung

(termasuk pengetahuan awal yang relevan). Sebagai KDD hasil proses, mungkin ada revisi dan perbaikan dari langkah ini. Memiliki memahami tujuan KDD, preprocessing data dimulai, sebagaimana didefinisikan dalam berikutnya tiga langkah (perhatikan bahwa beberapa metode di sini mirip dengan algoritma Data Mining, tetapi digunakan dalam konteks preprocessing).

2. **Memilih dan menciptakan** satu data set untuk mendukung proses penemuan knowledge akan dilakukan. Penentuan data yang akan digunakan untuk proses KDD dilakukan pada tahap ini. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data untuk KDD ke dalam sebuah data set, termasuk atribut yang diperlukan dalam proses KDD. Terdapat interaktif dan iteratif dari KDD tersebut. Dimulai dengan data yang tersedia baik mengatur dan kemudian mengembang dan mengamati efeknya dalam KDD.
3. **Preprocessing dan cleansing**. Dalam tahap ini kehandalan data ditingkatkan. Termasuk data clearing, seperti menangani data yang tidak lengkap, menghilangkan gangguan atau outlier. Termasuk mengunakan metode statistik yang kompleks, atau melakukan penambangan spesifik data dengan algoritma dalam KDD.
4. **Transformasi data**. Pada tahap ini, generasi data yang lebih baik untuk data mining dipersiapkan dan dikembangkan, membuat data menjadi lebih baik menggunakan metode reduksi dimensi dan transformasi atribut. Sebagai contoh, dalam pemeriksaan medis, hasil bagi atribut mungkin sering menjadi faktor yang paling penting, dan tidak satu persatu. Di pemasaran, kita mungkin perlu mempertimbangkan efek di luar kendali kita serta upaya dan temporal isu (seperti mempelajari pengaruh akumulasi iklan). Namun, bahkan jika kita tidak menggunakan transformasi yang tepat di awal, kita dapat memperoleh efek mengejutkan bahwa petunjuk kepada kita tentang transformasi diperlukan (Pada iterasi berikutnya). Dengan demikian proses KDD mencerminkan kepada dirinya sendiri dan menyebabkan pemahaman tentang transformasi yang dibutuhkan (seperti

pengetahuan ringkas dari sebuah ahli dalam bidang tertentu mengenai indikator terkemuka kunci).

5. **Memilih tugas data mining** yang cocok. Pada tahap ini ditentukan tipe data mining yang akan digunakan, apakah klasifikasi, regresi, atau clustering, tergantung pada tujuan KDD dan tahap sebelumnya.
6. **Memilih algoritma data mining**. Pemilihan algoritma yang paling tepat untuk menemukan pola dilakukan pada tahap ini. Ada dua tujuan utama dalam Data Mining: prediksi dan deskripsi. Prediksi sering disebut sebagai Supervised Data Mining, sementara deskriptif Data Mining meliputi aspek-aspek unSupervised dan visualisasi Data Mining. Sebagian besar data teknik pertambangan didasarkan pada pembelajaran induktif, di mana model yang dibangun secara eksplisit maupun implisit oleh generalisasi dari jumlah yang memadai pelatihan data training. Asumsi yang mendasari pendekatan induktif adalah bahwa model terlatih ini berlaku untuk kasus masa depan. Strategi ini juga memperhitungkan tingkat metalearning untuk set tertentu dari data yang tersedia.
7. **Penggunaan algoritma data mining**. Pada tahap ini dilakukan implementasi dari algoritma data mining yang telah ditentukan pada tahap sebelumnya. Misalnya dengan menseting parameter kontrol algoritma, seperti jumlah minimum kasus dalam daun tunggal dari pohon keputusan.
8. **Evaluasi**, pada tahap ini dilakukan evaluasi dan penerjemahan dari pola yang diperoleh, sehubungan dengan tujuan yang ditetapkan pada langkah pertama. Langkah ini berfokus pada komprehensibilitas dan kegunaan dari model induksi. Pada langkah ini pengetahuan ditemukan juga terdokumentasi untuk penggunaan lebih lanjut. Langkah terakhir adalah penggunaan dan umpan balik secara keseluruhan pada pola dan hasil penemuan diperoleh dengan Data Mining.
9. **Penggunaan pengetahuan yang ditemukan** yakni memasukkan pengetahuan ke dalam sistem lain untuk ditindaklanjuti. Pengetahuan menjadi aktif dalam arti bahwa kita dapat membuat perubahan ke sistem dan mengukur dampak Sebenarnya keberhasilan langkah ini menentukan efektivitas proses KDD secara keseluruhan. Ada banyak tantangan dalam

langkah ini, kehilangan "kondisi laboratorium". Misalnya, pengetahuan itu ditemukan dari sebuah snapshot statis tertentu (biasanya sampel) dari data, tapi sekarang data menjadi dinamis.

2.2.2. Pengelompokkan *Data Mining*

Pengelompokkan *data mining* dapat diklasifikasikan berdasarkan fungsi yang dilakukan atau berdasarkan jenis aplikasi yang menggunakannya [6].

1. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

Contoh lain klasifikasi dalam bisnis dan penelitian adalah :

a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
c. Mendiagnosa penyakit seorang pasien untuk mendapatkan termasuk kategori apa.

2. Pengklusteran (*Clustering*)

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain.Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian adalah :

a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
b. Untuk tujuan audit akutansi, yaitu melakukan pemisahan terhadap prilaku finansial dalam baik dan mencurigakan.

3. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Contoh asosiasi dalam bisnis dan penelitian adalah :

a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli bersamaan.

4. Deskripsi

Terkadang analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

5. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan dara sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan

level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

6. **Prediksi**

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.

2.3. KLASIFIKASI

Klasifikasi berperan untuk menemukan suatu bentuk model dan membedakan suatu kelas dari konsep data, setelah itu model digunakan untuk memprediksi label dari kelas objek yang status label kelasnya belum diketahui [7].

Klasifikasi juga dapat diartikan sebagai bentuk proses yang bertujuan untuk mencari dan menemukan model atau juga disebut fungsi yang berperan sebagai pembeda antara konsep atau kelas dari suatu objek yang tidak diketahui labelnya [8].

2.4. *DECISION TREE & ALGORITMA C4.5*

Menurut J. R. Quinlan [9], “A decision tree is a formalism for expressing such mappings. A tree is either a leaf node labeled with a class or a structure consisting of a test node linked to two or more subtrees. A test node computes some outcome based on the attribute values of an instance, where each possible outcome is associated with one of the subtrees”.

Algoritma C4.5 digunakan atau memiliki fungsi untuk menciptakan suatu pohon keputusan yang dimana pohon keputusan tersebut merupakan suatu metode klasifikasi dan juga prediksi yang banyak digunakan.

2.5. PENELITIAN SEJENIS

Tabel 2.1. Penelitian Sejenis

No.	Judul, Penulis dan Tahun	Metode	Atribut	Akurasi
1	"Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45". Randi Estian Pambudi, Sriyanto, Firmansyah. 2022. [1]	*Algoritma Decision Tree C4.5*	*Gender, Age, Hypertension, Heart Desease, Ever Married, Work Type, Avg Glukose Level, Bmi, Residence Status, Smoking Status, Stroke*	Hasil yang didapatkan untuk prediksi penyakit stroke sebesar 96.05%.
2	"Implementasi Algoritma Support Vector Machine (Svm) Untuk Prediksi Penyakit Stroke Dengan Atribut Berpengaruh". Ulfa Amelia, Jamaludin Indra, Anis Fitri Nur Masruriyah. 2022. [3]	*Algoritma Support Vector Machine.*	*Gender, Age, Hypertension, Heart Desease, Ever Married, Work Type, Residence Type, Average Glucose Level, BMI, Smoking Status.*	Hasil yang disimpulkan pada penelitian prediksi penyakit Strok menggunakan Algoritma SVM menghasilkan akurasi data sebesar 100%.
3	"Implementasi C45 Dalam Memprediksi Index Prestasi	*Algoritma C4.5*	Datang terlambat/telat, Ketepat waktu saat kumpul	Memperoleh tingkat persentase keakurasian

	Mahasiswa/i Menurut Kebiasaan Belajar". Mhd Yuda Rizki, Eva Sartika, Yoga Pratama, Suhaila Rasika, Agus Perdana Windarto. 2020. [10]		tugas, Bertanya saat berlangsungnya mata kuliah, Pemahaman dari seluruh mata kuliah/sems, Kehadiran.	sebesar 79,25%.
4	"Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke". Muhammad Naja Maskuri, Harliana, Kadek Sukerti, R.M. Herdian Bhakti. 2022. [11].	*Algoritma K-Nearest Neighbor.*	Jenis Kelamin, Umur, Hipertensi, Penyakit Jantung, Pernah Menikah, Tipe Pekerjaan, Tipe Tempat Tinggal, Rerata Kadar Glukosa, BMI, Status Merokok.	didapatkan nilai akurasi sebesar 95%.
5	"Implementasi Algoritma C.45 Untuk Klasifikasi Deteksi Serangan Pada Protokol	*Algoritma C4.5*	*Time, Source, Length, Protokol, Destination.*	Hasil analisis klasifikasi data jaringan dengan menggunakan menggunakan algoritma C.45 didapatkan nilai

	Jaringan". Irma Anggraeni, Siska Andriani. 2021. [12].			akurasi yaitu 93,62 %.
6	"Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke". Kenny Riva Sulaeman, Casi Setianingsih, Randy Erfa Saputra. 2022. [13].	*Algoritma Support Vector Machine.*	*Gender, Age, Hypertension, Heart Desease, Ever Married, Work Type, Residence Type, AVG Glucose Level, BMI, Smoking Status.*	Data unbalanced pada kernel linear mendapatkan akurasi terbesar 76% dan untuk polynomial mendapatkan hasil akurasi 80%. Untuk data yang balanced dengan kernel linear penulis mendapatkan hasil akurasi tertinggi 77%, sedangkan pada kernel polynomial penulis mendapatkan hasil tertinggi 76%.
7	"Prediction of Stroke using Data Mining Classification Techniques".	*C4.5 Algorithms*	*Class, Age, Gender, Number of Medication, Medication Name, Lab Test*	Based on this experiment the highest achieved accuracy was for C4.5 by 95.25%.

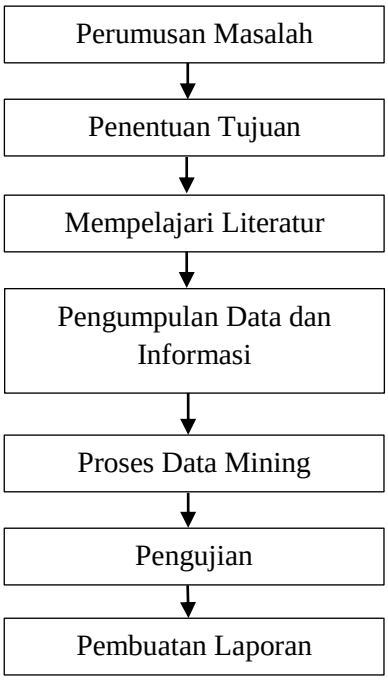
	Ohoud Almadani, Riyad Alshammari. 2018. [14].		*Name.*	
8	"Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online". Kiki Aidi Saputra, Jaya Tata Hardinata, Muhammad Ridwan Lubis, Sundari Retno Andani, Ilham Syahputra Saragih. 2020. [15].	*Algoritma C4.5*	*Tangible, Reliability, Responsiveness, Assurance, Dan Empathy.*	Dengan Algoritma C4.5. Menghasilkan 20 (dua puluh) rules dan Tingkat akurasi yang dihasilkan oleh metode tersebut adalah 86.67%.
9	"Penerapan Algoritma C4.5 Untuk Klasifikasi Penempatan Tenaga Marketing". Eka Fitriani, Riska	*Algoritma C4.5*	Jenis kelamin, Tinggi badan, Berat badan, Usia, Jarak dari tempat tinggal ke tempat kerja, Pengalaman, Tes kapabilitas,	Dapat disimpulkan bahwa nilai akurasi yang didapat model algoritma C4.5 adalah 91,10%.

	Aryanti, Atang Saepudin, Dian Ardiansyah. 2020. [16].		Targeting 3 bulan.	
10	"Implementasi Algoritma C4.5 Untuk Klasifikasi Tingkat Kepuasan Pembeli Online Shop". Dimas Bayu Febriyanto, Lovi Handoko, Wahyuli, Hanif Aisyah, Rumini. 2018. [17].	*Algoritma C4.5*	Produk, Pelayanan, Fasilitas.	Hasil penelitian terdapat tingkat akurasi sebesar 91%.

BAB III
METODOLOGI PENELITIAN

3.1. KERANGKA KERJA PENELITIAN

Untuk membantu dalam penyusunan penelitian ini, maka perlu adanya kerangka kerja (*framework*) yang jelas tahapan-tahapannya. Kerangka kerja ini merupakan langkah- langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Adapun kerangka kerja penelitian yang akan digunakan adalah sebagai berikut :



Gambar 3.1. Tahapan Penelitian

Berdasarkan kerangka kerja penelitian yang telah di gambar di atas, maka dapat di uraikan pembahasan masing masing tahapan dalam penelitian adalah sebagai berikut :

1. **Perumusan Masalah**

   Perumusan masalah pada penelitian ini adalah bagaimana penerapan data mining dengan *Algoritma C4.5* untuk memprediksi penyakit Stroke otak.

2. **Penentuan Tujuan**

   Tujuan yang akan dicapai dalam penelitian ini adalah ingin mengetahui berapa besar tingkat akurasi *Algoritma C4.5* dalam memprediksi Penyakit Stroke otak.

3. **Mempelajari Literatur**

   Mempelajari literatur-literatur yang dapat mencapai tujuan penelitian, literatur-literatur bersumber dari buku-buku perpustakaan Universitas Dinamika Bangsa Jambi dan jaringan internet. Literatur-literatur yang digunakan nanti dilampirkan dalam daftar pustaka.

4. **Pengumpulan Data dan Informasi**

   Dalam pengumpulan data, penulis mendapatkan dataset online yang terdapat pada sebuah website di internet.

5. **Proses Data Mining**

   Data Mining adalah proses pengektrasian *knowledge* yang tersimpan dalam dataset bervolume besar. Untuk mendapatkan *knowledge* dalam dataset menggunakan *Algoritma C4.5*.

6. **Pengujian**

   Pada tahap ini dilakukan pengujian dari hasil yang didapat dari tahap sebelumnya sebagai pedoman untuk mendapatkan hasil klasifikasi penyakit Stroke otak.

7. **Pembuatan Laporan**

   Pada tahap ini dilakukan pembuatan laporan, membuat hasil akhir dari suatu kegiatan penelitian berdasarkan data dan fakta yang telah diamati pada saat meneliti.

3.2. ANALISA TEKNIK *ALGORITMA C4.5*

Pada bagian ini data dan informasi yang diperoleh dan diproses dengan menggunakan *Algoritma C4.5* untuk mendapatkan hasil yang sesuai.

3.3. METODE PENELITIAN

Metode pengumpulan data yang digunakan dalam penelitian adalah observasi, dokumen, dan studi pustaka. Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Metode pengumpulan data dalam penelitian ini adalah sebagai berikut:

1. **Observasi**

   Merupakan metode pengumpulan data dengan cara melakukan pengamatan secara langsung pada objek yang diteliti yaitu pada kelulusan mahasiswa yang meliputi data kualitatif dan data kuantitatif dilakukan observasi untuk di analisis dengan Algoritma Naïve Bayes. Pengamatan data peninjauan ke Universitas Dinamika Bangsa Jambi untuk melihat dan mengetahui kondisi pembelajaran secara langsung sehingga mengetahui permasalahan yang terjadi.

2. **Dokumen**

   Pengambilan data melalui dokumen tertulis maupun elektronik dari lembaga atau institusi sesuai dengan yang diangkat dalam proposal ini yaitu di Universitas Dinamika Bangsa Jambi. Dokumen diperlukan untuk mendukung keperluan data yang lain.

3. **Studi Pustaka**

   Mengetahui informasi secara teori mengenai pokok permasalahan dan teori pendukung yang digunakan sebagai dasar pemikiran dalam membahas permasalahan yang ada, membaca referensi yang berhubungan dengan metode Naïve Bayes.

3.4. ALAT PENELITIAN

Penulis menggunakan beberapa alat/piranti yang digunakan untuk melakukan pengolahan data/bahan penelitian, yaitu:

1. Hardware, dengan spesifikasi sebagai berikut:
    a. Laptop, dengan processor Intel(R) Core i3 (2.00 GHz)
    b. RAM : 4.00 GB
    c. HDD : 1 TB
2. Software, dengan keterangan sebagai berikut:
    a. OS Windows 10 (64 bit)
    b. Microsoft Excel & Word 2016
    c. WEKA
    d. RapidMiner Studio Professional 7.1.001

3.5. JADWAL PENELITIAN

No	Jenis Kegiatan	Bulan																							
		Sept				Okt				Nov				Des				Jan				Feb			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Perumusan Masalah																								
2	Penentuan Tujuan																								
3	Mempelajari Literatur																								
4	Pengumpulan Data & Informasi																								
5	Proses Data Mining																								
6	Pengujian																								
7	Pembuatan Laporan																								

DAFTAR PUSTAKA

[1] R. E. Pambudi and F. I. Komputer, "Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45 1,2,3," vol. 16, no. x, pp. 221–226, 1978.

[2] Dr. Irwan SKM.M.Kes, *Epidemiologi Penyakit Menular*, vol. 109, no. 1. 2017.

[3] U. Amelia *et al.*, "Implementasi Algoritma Support Vector Machine (Svm) Untuk Prediksi Penyakit Stroke Dengan Atribut Berpengaruh," vol. III, pp. 254–259, 2022.

[4] E. Turban, J. E. Aronson, and T. P. Liang, "Decision Support Systems and Intelligent System,(Sistem Pendukung Keputusan dan Sistem Cerdas) Ed. 7. Jld. 2," 2005.

[5] L. Rokach and O. Maimon, "Clustering methods," in *Data mining and knowledge discovery handbook*, Springer, 2005, pp. 321–352.

[6] E. T. L. Kusrini and E. Taufiq, "Algoritma data mining," *Yogyakarta Andi Offset*, 2009.

[7] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.

[8] F. Aris, "Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi," vol. 1, no. 1, pp. 1–6, 2019.

[9] J. R. Quinlan, "Learning decision tree classifiers," *ACM Comput. Surv.*, vol. 28, no. 1, pp. 71–72, 1996, doi: 10.1145/234313.234346.

[10] M. Y. Rizki, E. Sartika, Y. Pratama, S. Rasika, A. P. Windarto, and A. D. A. N. Pembahasan, "Implementasi C45 Dalam Memprediksi Index Prestasi Mahasiswa / i Menurut Kebiasaan Belajar," pp. 297–303, 2020.

[11] M. N. Maskuri, K. Sukerti, and R. M. H. Bhakti, “Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke,” vol. 4, no. 1, pp. 130–140, 2022.

[12] I. Anggraeni and S. Andriani, “Implementasi Algoritma C.45 Untuk Klasifikasi Deteksi Serangan Pada Protokol Jaringan,” *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 18, no. 2, pp. 62–68, 2021, doi: 10.33751/komputasi.v18i2.3562.

[13] R. Sulaeman, C. Setianingsih, and R. Saputra, “Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke Support Vector Machine Algorithm Analysis In Stroke Disease Classification,” vol. 9, no. 3, pp. 922–928, 2022.

[14] O. Almadani and R. Alshammari, “Prediction of stroke using data mining classification techniques,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 457–460, 2018, doi: 10.14569/IJACSA.2018.090163.

[15] K. A. Saputra, J. T. Hardinata, M. R. Lubis, and S. R. Andani, “Klasifikasi Algoritma C4 . 5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online,” vol. 1, no. 3, pp. 113–118, 2020.

[16] E. Fitriani, R. Aryanti, A. Saepudin, and D. Ardiansyah, “Penerapan Algoritma C4.5 Untuk Klasifikasi Penempatan Tenaga Marketing,” *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 1, pp. 72–78, 2020, doi: 10.31294/p.v22i1.6898.

[17] D. Bayu Febriyanto, L. Handoko, and H. Aisyah, “Implementasi Algoritma C4.5 Untuk Klasifikasi Tingkat Kepuasan Pembeli Online Shop,” *J. Ris. Komput.*, vol. 5, no. 6, pp. 569–575, 2018, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C569

LAMPIRAN DATASET

gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
Female	71	0	0	Yes	Govt_job	Rural	193.94	22.4	smokes	1
Female	52	1	0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1
Female	79	0	0	Yes	Self-employed	Urban	228.7	26.6	never smoked	1
Male	82	0	1	Yes	Private	Rural	208.3	32.5	Unknown	1
Male	71	0	0	Yes	Private	Urban	102.87	27.2	formerly smoked	1
Male	80	0	0	Yes	Self-employed	Rural	104.12	23.5	never smoked	1
Female	65	0	0	Yes	Private	Rural	100.98	28.2	formerly smoked	1
Male	69	0	1	Yes	Self-employed	Urban	195.23	28.3	smokes	1
Male	57	1	0	Yes	Private	Urban	212.08	44.2	smokes	1
Male	42	0	0	Yes	Private	Rural	83.41	25.4	Unknown	1
Female	82	1	0	Yes	Self-employed	Urban	196.92	22.2	never smoked	1
Male	80	0	1	Yes	Self-employed	Urban	252.72	30.5	formerly smoked	1
Male	48	0	0	No	Govt_job	Urban	84.2	29.7	never smoked	1
Female	82	1	1	No	Private	Rural	84.03	26.5	formerly smoked	1
Male	74	0	0	Yes	Private	Rural	219.72	33.7	formerly smoked	1
Female	72	1	0	Yes	Private	Rural	74.63	23.1	formerly smoked	1
Male	58	0	0	No	Private	Rural	92.62	32	Unknown	1
Female	49	0	0	Yes	Private	Urban	60.91	29.9	never smoked	1
Male	78	0	0	Yes	Private	Rural	78.03	23.9	formerly smoked	1
Male	54	0	0	Yes	Private	Urban	71.22	28.5	never smoked	1
Male	82	0	1	Yes	Private	Urban	144.9	26.4	smokes	1
Male	60	1	0	Yes	Govt_job	Urban	213.03	20.2	smokes	1
Male	76	1	0	Yes	Private	Rural	243.58	33.6	never smoked	1
Female	58	0	0	Yes	Private	Urban	107.26	38.6	formerly smoked	1
Male	81	0	0	Yes	Self-employed	Urban	99.33	33.7	never smoked	1
Female	39	1	0	Yes	Private	Rural	58.09	39.2	smokes	1
Female	79	0	1	Yes	Private	Rural	127.29	27.7	never smoked	1
Female	77	1	0	Yes	Self-employed	Urban	124.13	31.4	never smoked	1
Male	63	0	1	Yes	Private	Rural	196.71	36.5	formerly smoked	1
Female	82	0	0	Yes	Private	Rural	59.32	33.2	never smoked	1
Male	73	1	0	Yes	Self-employed	Urban	194.99	32.8	never smoked	1
Female	54	1	0	Yes	Govt_job	Urban	180.93	27.7	never smoked	1
....	....	....	....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....	....	....	....
Female	45	1	0	Yes	Govt_job	Rural	95.02	31.8	smokes	0
Male	40	0	0	Yes	Private	Rural	83.94	30	smokes	0
Female	80	1	0	Yes	Private	Urban	83.75	29.1	never smoked	0