

ANALISIS PERBANDINGAN HASIL DARI ALGORITMA DECISION TREE C4.5 MENGGUNAKAN MICROSOFT EXCEL DAN CHATGPT

TUGAS AKHIR

Disusun sebagai salah satu syarat untuk mengambil Kualifikasi Penelitian di Program Studi
Teknik Informatika Universitas Pasundan Bandung

oleh :

Try Utami Kemala Ningrum

NRP : 18.304.0125



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PASUNDAN
SEPTEMBER 2025**

LEMBAR PENGESAHAN
LAPORAN TUGAS AKHIR

Telahn diujikan dan dipertahankan dalam Sidang Sarjana Program Studi Teknik Informatika Universitas Pasundan Bandung, pada hari ini dan tanggal sidang sesuai berita acara sidang. Tugas akhir dari :

Nama : Try Utami Kemala Ningrum

NRP : 18.304.0125

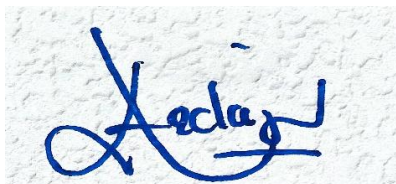
Dengan judul :

**“ANALISIS PERBANDINGAN HASIL DARI ALGORITMA DECISION TREE C4.5
MENGUNAKAN MICROSOFT EXCEL DAN CHATGPT”**

Bandung, 29 September 2025

Mengatahui,

Pembimbing Utama



(Dr. Ririn Dwi Agustin, ST. MT.)

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa :

1. Tugas akhir ini adalah benar-benar asli dan belum pernah diajukan untuk mendapatkan gelar akademik, baik di Universitas Pasundan Bandung maupun di perguruan tinggi lainnya.
2. Tugas akhir ini merupakan gagasan, rumusan dan penelitian saya sendiri, tanpa bantuan pihak lain kecuali arahan dari tim dosen Pembimbing.
3. Dalam tugas akhir ini tidak terdapat karya atau pendapat orang lain, kecuali bagian-bagian tertentu dalam penulisan laporan Tugas Akhir yang saya kutip dari hasil karya orang lain telah dituliskan dalam sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan karya ilmiah, sert disebutkan dalam Daftar Pustaka pada tugas akhir ini.
4. Kakas, perangkat lunak, dan alat bantu kerja lainnya yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab saya, bukan tanggung jawab Universitas Pasundan Bandung.

Apabila di kemudian hari ditemukan seluruh atau sebagian laporan tugas akhir ini bukan hasil karya saya sendiri atau adanya plagiasi dalam bagian-bagian tertentu, saya bersedia menerima sanksi akademik, termasuk pencabutan gelar akademik yang saya sandang sesuai dengan norma yang berlaku di Universitas Pasundan Bandung, serta perundangan-undangan lainnya.

Bandung, 29 September 2025

Yang membuat pernyataan



(Try Utami Kemala Ningrum)

Nrp. 18.304.0125

ABSTRAK

Algoritma Decision Tree C4.5 merupakan salah satu metode klasifikasi yang banyak digunakan karena kemampuannya dalam menghasilkan model yang mudah dioahami serta mampu menangani data dengan atribut numerik maupun kategorik. Penelitian ini bertujuan untuk menganalisis dan membandingkan hasil penerapan algoritma C4.5 menggunakan dua pendekatan yaitu perhitungan manual dengan bantuan Microsoft Excel serta pemanfaatan ChatGPT sebagai alat komputasi berbasis kecerdasan buatan.

Proses analisis dilakukan melalui tahap perhitungan nilai entropy, information gain, hingga pembentukan pohon keputusan. Hasil penelitian menunjukan bahwa kedua pendekatan menghasilkan struktur decision tree yang sejalan, meskipun terdapat perbedaan pada kecepatan, efisiensi, serta kemudahan dlam interpretasi hasil. Penggunaan Microsoft Excel memberikan pemahaman mendalam mengenai proses perhitungan setiap langkah algoritma, sementara ChatGPT menawarkan efisiensi waktu dan fleksibilitas dalam eksplorasi kasus klasifikasi.

Dengan demikian, kombinasi keduanya dapat menjadi alternatif metode pembelajaran sekaligus alat analisis yang saling melengkapi dalam memahami penerapan algoritma C4.5.

Kata Kunci : Decision Tree C4.5, Microsoft Excel, ChatGPT, Klasifikasi.

ABSTRACT

The Decision Tree C4.5 algorithm is one of the most widely used classification methods due to its ability to generate models that are easy to interpret while handling both numerical and categorical attributes. This study aims to analyze and compare the result of applying the C4.5 algorithm using two different approaches : manual calculation with the assistance of Microsoft Excel and the utilization of ChatGPT as an artificial intelligence-based computational tool.

The analysis process includes calculating entropy, information gain, and constructing the decision tree. The findings show that both approaches produce consistent decision tree structures, although differences are observed in terms of speed, efficiency, and ease of interpretation. Microsoft Excel provides a deeper understanding of each calculation step within the algorithm, whereas ChatGPT offers time efficiency and flexibility in exploring classification cases.

Therefore, the combination of both approaches can serve as an alternative method for learning as well as a complementary analytical tool in understanding the implementation of the C4.5 algorithm.

Keywords : Decision Tree C4.5, Microsoft Excel, ChatGPT, Classification.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunianya sehingga penulis dapat menyelesaikan salah satu syarat memperoleh gelar sarjana Teknik Informatika di Universitas Pasundan Bandung.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna seperti yang diharapkan. Maka pada kesempatan ini penulis sampaikan terima kasih sebesar-besarnya atas segala bantuan yang penulis terima baik secara moral maupun material, sehingga penulis dapat menyelesaikan laporan ini kepada :

1. Kepada pembimbing utama, ibu Dr. Ririn Dwi Agustin, ST. MT yang telah memberikan nasehat dan arahan kepada penulis dalam menyelesaikan Tugas Akhir.
2. Kepada bapak Rudi Hendrawan yang senantiasa memberikan arahan dengan baik untuk kelancaran Tugas Akhir.
3. Kepada Orang Tua Tersayang, dan keluarga yang telah selalu memberikan motivasi serta doanya dalam pembuatan Tugas Akhir ini.
4. Koordinator Tugas Akhir serta seluruh civitas akademik Teknik Informatika di Universitas Pasundan, yang telah memberikan bekal selama menimba ilmu.
5. Kepada teman-teman seperjuangan Universitas Pasundan yang tidak bisa semua penulis sebutkan.

Tiada gading yang tak retak, tiada gelombang tanpa ombak segala kesalahan menyusun tugas akhir ini tentu saja penulis mendapatkan kesulitan dan hambatan. Oleh karena itu penulis harapkan kritik dan saran dari semua pihak demi perbaikan dimasa yang akan datang.

Dengan ikhlas penulis dapat memanjatkan doa kepada Tuhan Yang Maha Esa agar dapat membalas jasa Bapak/Ibu serta rekan – rekan sekalian.

Bandung, 29 September 2025



Try Utami Kemala Ningrum

DAFTAR ISI

LEMBAR PENGESAHAN	i
LEMBAR PERNYATAAN KEASLIAN	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
BAB 1 PENDAHULUAN	1-1
1.1 Latar Belakang	1-1
1.2 Rumusan Masalah	1-3
1.3 Tujuan Penelitian Tugas Akhir	1-3
1.4 Manfaat Penelitian Tugas Akhir	1-3
1.5 Lingkup Tugas Akhir	1-4
1.6 Sistematika Penulisan Tugas Akhir	1-4
BAB 2 LANDASAN TEORI DAN PENELITIAN TERDAHULU.....	2-5
2.1 Machine Learning	2-5
2.1.1 Definisi dan Konsep Dasar	2-5
2.1.2 Jenis-Jenis Algoritma Machine Learning.....	2-5
2.1.3 Posisi Decision Tree dalam Machine Learning.....	2-6
2.2 Decision Tree	2-7
2.2.1 Pengertian Decision Tree	2-7
2.2.2 Konsep Dasar & Struktur Decision Tree.....	2-7
2.2.3 Fungsi dalam Klasifikasi Data	2-8

2.2.4 Kelebihan dan Kekurangan	2-8
2.3 Algoritma C4.5.....	2-9
2.3.1 Pengertian Algoritma C4.5.....	2-9
2.3.2 Kinerja Algoritma C4.5 dalam Berbagai Platform.....	2-10
2.3.3 Implementasi dan Hasil dengan ChatGPT	2-11
2.3.4 Kelebihan dan Kekurangan C4.5	2-11
2.4 Perkembangan ChatGPT, GPT Analytic, Buatan OpenAI	2-13
2.4.1 GPT Analytic.....	2-14
2.5 Penelitian Terdahulu	2-15
BAB 3	2-17
METODOLOGI PENELITIAN.....	3-18
3.1 Alur Penelitian.....	3-18
3.2 Perumusan Masalah.....	3-19
3.2.1 Analisis Sebab Akibat	3-19
3.2.2 Solusi Masalah	3-21
3.3 Memahami Algoritma Decision Tree (ID3 atau C4.5)	3-24
3.4 Simulasi Perhitungan Menggunakan Excel Untuk 2 Contoh Data	3-25
3.5 Melakukan Prompting ke GPT Data Analytic	3-27
3.6 Kesimpulan Tingkat Validasi dari ChatGPT Data Analytic	3-28
BAB 4 HASIL DAN PEMBAHASAN.....	4-30
4.1 Deskripsi Dataset Penelitian.....	4-30
4.2 Hasil Simulasi Perhitungan di Microsoft Excel	4-32
4.3 Hasil Simulasi Perhitungan di ChatGPT Data Analytic.....	4-48
4.4 Perbandingan Hasil Perhitungan	4-64
4.5 Analisis Pengalaman Mahasiswa	4-65
BAB 5	5-67

KESIMPULAN DAN SARAN.....	5-67
5.1 KESIMPULAN	5-67
5.2 SARAN	5-68
DAFTAR PUSTAKA	5-69

DAFTAR TABEL

Table 2.1 Penelitian Terdahulu.....	2-15
Table 4.1 Perbandingan Hasil Perhitungan Dataset Mawar.	4-64
Table 4.2 Perbandingan Hasil Perhitungan Dataset Resiko Diabetes.	4-65
Table 4.3 Ringkasan Hasil Kuesioner Pengalaman Mahasiswa.....	4-65

DAFTAR GAMBAR

Gambar 3 1 Alur Penelitian.....	3-18
Gambar 3 2 Fishbone Masalah.....	3-20
Gambar 3 3 Fishbone Solusi	3-22
Gambar 4 1 Dataset Mawar	4-30
Gambar 4 2 Dataset Resiko Diabetes.....	4-31
Gambar 4 3 Hitung Entropy Mawar	4-32
Gambar 4 4 Hitung Information Gain Tiap Atribut	4-33
Gambar 4 5 Hitung Information Gain Terbesar	4-33
Gambar 4 6 Dataset Kelembapan Tanah Sedang	4-35
Gambar 4 7 Hitung Entropy Kelembapan Tanah Sedang	4-35
Gambar 4 8 Hitung Information Gain Tiap Atribut Kelembapan Tanah Sedang	4-36
Gambar 4 9 Hitung Information Gain Terbesar Kelembapan Tanah Sedang.....	4-36
Gambar 4 10 Dataset Warna Kelopak Merah.....	4-38
Gambar 4 11 Perhitungan Keseluruhan Warna Kelopak Merah	4-38
Gambar 4 12 Dataset Warna Kelopak Pink.....	4-39
Gambar 4 13 Perhitungan Keseluruhan Warna Kelopak Pink	4-40
Gambar 4 14 Pohon Keputusan Mawar	4-41

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Latar belakang penelitian ini berfokus pada analisis perbandingan hasil dari algoritma Decision Tree C4.5, menggunakan Microsoft Excel dan ChatGPT. Pemanfaatan AI dalam konteks pendidikan dan analisis data dapat dilihat dalam penelitian yang menyoroti potensi kecerdasan buatan dalam mendukung pembelajaran dan pemahaman siswa terhadap materi kompleks, termasuk analisis data dengan menggunakan Decision Tree. Microsoft Excel sebagai alat, mahasiswa dapat langsung terlibat dalam proses perhitungan dan visualisasi data yang mendukung pemahaman. Selain itu, ChatGPT memungkinkan interaksi yang dinamis, memberikan ruang bagi mahasiswa untuk eksplorasi mandiri dan pemecahan masalah yang lebih mendalam.

Machine learning, sebagai salah satu cabang dari kecerdasan buatan, telah menjadi fokus perhatian luas. Salah satu algoritma yang paling dikenal dalam machine learning adalah Decision Tree C4.5, yang digunakan untuk mengklasifikasi dan analisis data. Algoritma ini diharapkan mampu memberikan hasil yang akurat, memungkinkan peneliti dan profesional untuk membuat keputusan berdasarkan data yang kompleks. Keakuratan hasil yang dihasilkan oleh Decision Tree sangat dipengaruhi oleh pemilihan fitur dan metode pengolahan data yang digunakan dalam pembangunan model. Dalam konteks ini, pemanfaatan Excel sebagai alat analisis memungkinkan pemodelan yang terstruktur dan sistematis, di mana pengguna dapat dengan mudah memahami langkah-langkah dari input hingga output.

Predictive modeling yang diterapkan oleh algoritma Decision Tree, termasuk C4.5, juga mengindikasikan adanya metode yang sistematis dalam mengatasi masalah klasifikasi. Kekuatan utama dari Decision Tree terletak pada kemampuannya untuk menetapkan aturan yang jelas yang dapat digunakan untuk memprediksi kelas dari data baru. Di sisi lain, ChatGPT menawarkan pendekatan yang lebih interaktif, memberikan siswa kemampuan untuk menyelidiki dan menganalisis data dengan cara yang lebih konversasional. Pendekatan ini tidak hanya mempercepat proses belajar, tetapi juga

meningkatkan pemahaman konsep-konsep yang lebih kompleks, termasuk di dalamnya algoritma seperti C4.5.

Dalam penelitian ini, peneliti mencoba untuk menunjukkan cara belajar algoritma Decision Tree, dan membandingkan hasil kalkulasi akurat yang diberikan oleh pengajar di aplikasi Excel dengan hasil yang diperoleh melalui interaksi mandiri mahasiswa di platform ChatGPT. Melalui pendekatan ini, diharapkan mahasiswa dapat merasakan langsung perbedaan antara analisis berbasis manual dan berbasis teknologi AI, di mana masing-masing memiliki kelebihan dan batasannya sendiri. Sebagai contoh, keputusan yang diambil berdasarkan output dari Excel mungkin lebih terstruktur, sedangkan hasil dari ChatGPT dapat mencakup pemahaman berdasarkan konteks.

Pendekatan pembelajaran ini juga didukung oleh penelitian yang menunjukkan bahwa kinerja analisis data menggunakan Decision Tree dapat bervariasi tergantung pada kompleksitas dataset dan keahlian pengguna dalam mengimplementasikan algoritma tersebut. Dalam konteks ini, pembelajaran berbasis AI tidak hanya menawarkan hasil yang nyata dan relevan, tetapi juga menyediakan pengalaman pembelajaran yang berharga bagi mahasiswa dalam memahami penggunaan algoritma dalam praktik nyata. Hasil perbandingan yang dihasilkan akan memicu diskusi lebih lanjut mengenai kelebihan dan kekurangan masing-masing metode, serta penerapan praktis dari algoritma Decision Tree dalam berbagai konteks.

Konsep keputusan berbasis data menjadi sangat penting dalam dunia modern di mana data memainkan peran sentral dalam pengambilan keputusan. Seiring dengan kemajuan teknologi, integrasi alat analisis seperti Excel dengan kekuatan pemodelan dari AI seperti ChatGPT memberikan jalan baru bagi mahasiswa untuk mengembangkan keterampilan analitis dan pemecahan masalah. Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi penting dalam literatur mengenai pembelajaran berbasis alat AI dan analisis data, serta bagaimana kedua pendekatan tersebut saling melengkapi dalam konteks pendidikan.

Akhirnya, penelitian ini tidak hanya bertujuan untuk mengevaluasi efektivitas masing-masing metode, tetapi juga untuk menyelidiki dampak dari penggunaan teknologi AI dalam proses belajar mengajar. Hal ini sejalan dengan penelitian-penelitian sebelumnya yang menyebutkan bahwa penggunaan alat berbasis AI dapat meningkatkan

keterlibatan siswa, menciptakan pengalaman belajar yang lebih interaktif, dan memfasilitasi pemahaman lebih dalam terhadap konsep-konsep yang diajarkan. Dengan demikian, hasil dari penelitian ini diharapkan mampu memberikan wawasan yang lebih luas kepada pendidik dan praktisi mengenai efektivitas penggunaan algoritma Decision Tree C4.5 dalam konteks pembelajaran dengan hasil kalkulasi yang akurat dan interaktif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimanakah mekanisme kerja Algoritma Decision Tree C4.5 yang dideskripsikan menggunakan Excel untuk beberapa contoh studi kasus?
2. Bagaimana alur Prompt di ChatGPT AI untuk mendapatkan hasil perhitungan seperti halnya hasil eksekusi Algoritma di Excel?
3. Bagaimana pengalaman beberapa mahasiswa dalam memahami Algoritma decision Tree C4.5 setelah berinteraksi dengan ChatGPT AI?

1.3 Tujuan Penelitian Tugas Akhir

Tujuan dari Tugas Akhir ini adalah untuk mendeskripsikan mekanisme kerja Algoritma decision Tree C4.5 melalui contoh studi kasus yang diimplementasikan di Excel, merancang alur prompt yang optimal untuk mendapatkan hasil perhitungan menggunakan ChatGPT, serta menganalisis tingkat pemahaman mahasiswa terhadap algoritma Decision Tree C4.5 dengan membandingkan pendekatan tradisional dari basis AI.

1.4 Manfaat Penelitian Tugas Akhir

Penelitian ini diharapkan dapat memberikan manfaat, baik secara teoritis maupun praktis, sebagai berikut :

1. Teoritis : Memberikan kontribusi terhadap perkembangan ilmu pengetahuan dalam bidang *data mining*, khususnya dalam penerapan algoritma C4.5 menggunakan dua pendekatan yang berbeda, yaitu Microsoft Excel dan ChatGPT.
2. Praktis : Memberikan alternatif metode bagi pengguna yang membutuhkan solusi praktis dalam penerapan algoritma C4.5 tanpa perlu perangkat lunak khusus yang

rumit dan mahal. Hasil penelitian ini juga diharapkan dapat digunakan oleh praktisi dalam memilih alat yang paling sesuai untuk analisis data dengan algoritma C4.5.

1.5 Lingkup Tugas Akhir

Penelitian ini dibatasi pada :

1. Studi kasus sederhana yang mencakup data klasifikasi.
2. Penggunaan Excel untuk simulasi manual algoritma Decision Tree C4.5.
3. Interaksi mahasiswa dengan ChatGPT dalam melakukan analisis data menggunakan model Data Analytic.
4. Pengumpulan dan analisis data melalui kuesioner terkait pengalaman mahasiswa.

1.6 Sistematika Penulisan Tugas Akhir

Penulisan skripsi ini disusun dengan sistematika sebagai berikut :

- **Bab I : Pendahuluan**
Berisi latar belakang rumusan masalah, tujuan, manfaat, ruang lingkup penelitian dan sistematika penulisan tugas akhir.
- **Bab II : Tinjauan Pustaka**
Menyajikan kaitan teori terkait dengan algoritma C4.5, Decision Tree, Microsoft Excel, ChatGPT, dan penelitian-penelitian terdahulu yang relevan.
- **Bab III : Metodologi Penelitian**
Menjelaskan metode penelitian yang digunakan termasuk desain eksperimen pemilihan dataset, serta cara implementasi dan pengolahan data.
- **Bab IV : Hasil dan Pembahasan**
Menyajikan hasil eksperimen dan analisis terhadap perbandingan penggunaan Excel dan ChatGPT dalam implementasi C4.5
- **Bab V : Kesimpulan dan Saran**
Menyimpulkan hasil penelitian dan memberikan saran untuk penelitian lebih lanjut.

BAB 2

LANDASAN TEORI DAN PENELITIAN TERDAHULU

2.1 Machine Learning

2.1.1 Definisi dan Konsep Dasar

Machine learning (ML) adalah cabang dari kecerdasan buatan (Artificial Intelligence/AI) yang memungkinkan komputer belajar dari data tanpa harus diprogram secara eksplisit. Dengan kata lain, Machine Learning membuat sistem dapat mengenali pola, membuat prediksi, atau mengambil keputusan berdasarkan data yang telah ada.

Konsep Dasar Machine Learning :

1. Data : Input berupa data (angka, teks, gambar, suara, dan lain sebagainya).
2. Model : Refersentasi matematis/statistik yang mempelajari hubungan dari data.
3. Training : Proses “belajar” dari data dengan cara menemukan pola.
4. Testing/Evaluasi : Mengukur seberapa baik model bekerja pada data baru.
5. Prediksi : Menghasilkan output/keputusan berdasarkan input baru.

2.1.2 Jenis-Jenis Algoritma Machine Learning

Secara umum, Algoritma Machine Learning dibagi menjadi 3 kategori besar :

a. Supervised Learning

- Definisi : Algoritma dilatih dengan data berlabel (ada input dan output yang benar).
- Tujuan : Memprediksi output yang baru.
- Contoh Algoritma :
 - Linear Regression.
 - Logistic Regression.
 - Decision Tree.
 - Random Forest.
 - Support Vector Machine (SVM).
 - Neyral Network.

b. Unsupervised Learning

- Definisi : Algoritma belajar dari data tanpa label (tidak ada jawaban benar yang diberikan).
- Tujuan : Menemukan pola, struktur, atau pengelompokan data.

- Contoh Algoritma :
 - K-Means Clustering.
 - Hierarchical Clustering.
 - Principal Component Analysis (PCA).
 - Autoencoder.

c. Reinforcement Learning

- Definisi : Algoritma belajar melalui interaksi dengan lingkungan, menggunakan sistem reward (penghargaan) dan penalty (hukuman).
- Tujuan : Membuat agen mengambil keputusan optimal berdasarkan pengalaman.
- Contoh Aplikasi :
 - Robotik.
 - Game AI (misalnya AlphaGo).
 - Sistem rekomendasi dinamis.

2.1.3 Posisi Decision Tree dalam Machine Learning

Posisi Decision Tree dalam Machine learning dapat kita lihat sebagai berikut :

- Decision Tree termasuk kedalam algoritma Supervised Learning.
- Jenis tugas yang bisa diselesaikan :
 - Klasifikasi.
Contoh : Menentukan apakah email adalah spam atau tidak spam.
 - Regresi.
Contoh : Memprediksi harga rumah berdasarkan atribut tertentu.
- Kelebihan Decision Tree :
 - Mudah dipahami (mirip dengan logika manusia dalam membuat keputusan).
 - Bisa menangani data kategorikal maupun numerik.
 - Tidak memerlukan banyak preprocessing data.
- Kekurangan :
 - Rentan terhadap overfitting.
 - Tidak stabil jika terjadi perubahan kecil pada data (solusinya biasanya menggunakan Random Forest atau Boosting).

2.2 Decision Tree

2.2.1 Pengertian Decision Tree

Decision Tree (pohon keputusan) merupakan metode supervised learning non-parametrik yang digunakan untuk mengklasifikasi maupun regresi. Pohon ini memprediksi nilai variabel target menggunakan aturan keputusan sederhana yang disimpulkan dari data secara visual, berbentuk struktur pohon dengan node dan cabang.

Dalam konteks klasifikasi, decision tree disebut juga sebagai classification tree, yaitu merupakan model yang memprediksi hasil berupa kategori (kualitatif) melalui pemisahan data berulang berdasarkan fitur.

2.2.2 Konsep Dasar & Struktur Decision Tree

Struktur dasar ini terdiri dari :

- Root Node :
Merupakan simpul akar yang mewakili keseluruhan dataset. Dari sinilah langkah klasifikasi dimulai.
- Internal Nodes (Decision Nodes) :
Menyatakan pengujian atau aturan berdasarkan fitur pada data.
- Branches (Cabang) :
Merupakan jalur yang menghubungkan node setelah pengujian atribut.
- Leaf Nodes (Daun) :
Merupakan simpul akhir yang menunjukkan hasil keputusan/kategori.

Proses Pembentukan Decision Tree

1. Mulai dari root node (seluruh dataset).
2. Pilih atribut terbaik, misalnya menggunakan information gain, gini impurity, atau gain ratio untuk membagi data.
3. Bagi dataset berdasarkan nilai atribut tersebut dan buat simpul decision tree yang sesuai.
4. Ulangi secara rekursif hingga mencapai kondisi berhenti :
 - Semua data di node memiliki kelas yang sama.
 - Tidak adanya fitur tersisa.
 - Tidak ada peningkatan kualitas prediksi.

2.2.3 Fungsi dalam Klasifikasi Data

Adapun fungsi dalam klasifikasi sebagai berikut :

- **Pembagian Kategori.**
Decision tree memisahkan data kedalam kelas-kelas berdasarkan fitur, efektif untuk variabel target kategori (contoh : ya/tidak).
- **Referensi Visual & Interpretasi Mudah.**
Strukturnya intuitif, memudahkan interpretasi dan komunikasi hasil, bahkan oleh yang non-teknis.
- **Pemrosesan Minimal.**
Tidak memerlukan banyak pra-pemrosesan seperti normalisasi juga mampu menangani data numerik maupun kategorikal secara langsung, termasuk nilai hilang.
- **Penanganan Data Non-Linear & Fitur Kompleks.**
Dapat mengungkap pola-pola kompleks dalam data yang sulit diraba oleh metode lain.
- **Fleksibilitas dan Efisiensi.**
Setelah dibangun, prediksi cepat dilakukan hanya dengan traversing dari akar ke daun.

2.2.4 Kelebihan dan Kekurangan

- **Kelebihan.**
 - Mudah dipahami dan divisualisasikan, cocok untuk interpretasi manusia.
 - Tidak banyak pra-pemrosesan data dan bisa menangani fitur numerik/kategorikal serta missing values.
 - Dapat menangani pola non-linear dan fleksibel dalam berbagai jenis data.
 - Cukup efisien dalam prediksi terutama setelah struktur pohon terbentuk.
- **Kekurangan.**
 - Cenderung overfitting, terutama jika pohon terlalu dalam atau kompleks.
 - Sensitif terhadap perubahan kecil dalam data struktur pohon bisa berubah secara drastis.

- Bias terhadap atribut dengan banyaknya kategori, yang tidak selalu paling informatif.

2.3 Algoritma C4.5

2.3.1 Pengertian Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang sering digunakan didalam data mining untuk membangun sebuah model pohon keputusan. Ross Quinlan pada tahun 1993 menyebutkan bahwa algoritma C4.5 merupakan hasil dari algoritma sebelumnya yang dikenal sebagai ID3 (Iterative Dichotomiser 3). Algoritma C4.5 merupakan salah satu algoritma populer dalam klasifikasi data yang dikembangkan oleh Ross Quinlan pada tahun 1993. Algoritma C4.5 adalah penyempurnaan dari algoritma ID3 sebelumnya, dengan kemampuan untuk menangani atribut kontinu, mengatasi masalah nilai yang hilang, dan menghasilkan pohon keputusan yang lebih efisien. Salah satu fitur utama dari algoritma ini adalah penggunaan konsep "gain ratio" untuk menentukan atribut yang paling relevan dalam membagi dataset menjadi subset yang lebih homogen dalam konteks kelas target.

Algoritma C4.5 memilih atribut terbaik untuk menjadi pertanyaan pada setiap cabang pohon, sehingga Keputusan diambil dengan langkah-langkah yang efisien dan akurat. Berikut langkah-langkah pengerjaan algoritma C4.5 dari awal sampai terbentuknya pohon :

1. Siapkan Dataset
 - Berisi data training dengan atribut (fitur) dan label kelas.
 - Contoh : Data Bunga dengan atribut warna, ukuran, kelembapan, dan label jenis .
2. Hitung Entropy Dataset (Data Awal)
 - Rumus :

$$Entropy(S) = \sum_{i=1}^k -p_i \log_2 p_i$$
 - p_i = propori kelas ke-i.
 - Entropy menunjukan tingkat ketidakpastian :
 - 0 = Semua data 1 kelas (pasti).
 - 1 = Data merata di semua kelas (tidak pasti).
3. Hitung Entropy untuk Tiap Atribut
 - Bagi data berdasarkan nilai atribut.
 - Hitung entropy masing-masing subset.

- Lalu hitung rata-ratanya (weighted average).
- 4. Hitung Information Gain (IG)
 - Rumus :

$$Gain(S,A) = Entropi(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropi(S_i)$$
 - Pilih atribut dengan nilai gain terbesar sebagai node (simpul) keputusan.
- 5. Buat Cabang Pohon
 - Setiap nilai dari atribut menjadi cabang.
 - Ulangi proses *entropy – gain – pilih atribut* terbaik untuk setiap cabang.
- 6. Berhenti jika :
 - Semua data dalam cabang memiliki kelas yang sama.
 - Tidak ada atribut yang tersisa.
 - Atau data kosong (gunakan mayoritas kelas sebelumnya).
- 7. Hasil Pohon Keputusan
 - Pohon dapat dikonversi menjadi aturan IF-THEN untuk memudahkan implementasi.

2.3.2 Kinerja Algoritma C4.5 dalam Berbagai Platform

Dalam aplikasi praktisnya, algoritma C4.5 dapat diimplementasikan menggunakan berbagai platform, salah satunya adalah Microsoft Excel, yang memungkinkan analisis data yang relatif sederhana dan mudah dipahami. Excel menyediakan fungsi dasar yang dapat membantu dalam menerapkan logika dasar dari algoritma C4.5 meskipun terbatas dalam hal kompleksitas dan performa ketika dibandingkan dengan alat khusus. Kinerja algoritma C4.5 pada Excel dapat menghasilkan tingkat akurasi yang kompetitif, tetapi sering kali tidak maksimal dalam pemrosesan dataset yang sangat besar dibandingkan dengan implementasi di lingkungan pemrograman yang lebih canggih seperti Python atau R (Suherman et al., 2020).

Sebuah studi oleh Suherman et al. menunjukkan bahwa penggunaan C4.5 dalam konteks klasifikasi hasil ujian menggunakan pendekatan optimasi swarm partikel menghasilkan akurasi hingga 97.13% (Suherman et al., 2020). Ini menunjukkan bahwa meskipun Excel memiliki keterbatasan dalam pengaturan data yang besar, ia masih dapat memberikan hasil yang cukup baik dalam pengaplikasian algoritma C4.5.

2.3.3 Implementasi dan Hasil dengan ChatGPT

Sebaliknya, implementasi algoritma C4.5 melalui antarmuka seperti ChatGPT dapat memberikan kemudahan dalam hal interaksi pengguna. Dengan kemampuannya untuk membaca dan memproses berbagai data secara cepat, ChatGPT dapat dioptimalkan untuk menjalankan algoritma C4.5 dengan menggunakan backend yang lebih kuat, seperti server berbasis Python yang dijalankan secara otomatis. Hal ini memungkinkan analisis yang lebih mendalam dan akurasi prediksi yang lebih tinggi dibandingkan dengan Excel.

Menurut Liu dan Sun, eksperimen yang membandingkan beberapa algoritma decision tree, termasuk C4.5, menemukan bahwa C4.5 memiliki waktu eksekusi yang kompetitif bahkan ketika dataset mencapai ukuran yang sangat besar, menunjukkan efisiensi yang baik (Liu & Sun, 2022). Penggunaan ChatGPT untuk menerapkan algoritma ini berpotensi meningkatkan kecepatan dan akurasi, khususnya dalam situasi di mana analisis real-time diperlukan.

Perbandingan antara implementasi C4.5 di Microsoft Excel dan ChatGPT menunjukkan bahwa meskipun kedua metode dapat menghasilkan hasil yang akurat, masing-masing memiliki kelebihan dan kekurangan. Excel lebih cocok untuk analisis sederhana dan interaksi manual, sementara ChatGPT dapat secara dramatis meningkatkan efisiensi dan kemampuan analitis melalui otomatisasi dan analisis berbasis AI. Keduanya dapat saling melengkapi tergantung pada konteks dan kebutuhan penggunaannya.

Analisis yang lebih mendalam dapat dilakukan dengan memperluas studi ke algoritma lain yang sebanding, seperti RIPPER, yang juga merupakan algoritma klasifikasi populer. Dalam beberapa konteks, C4.5 menunjukkan performa yang lebih baik dalam hal akurasi dibandingkan dengan algoritma lain, sehingga menegaskan kekuatannya dalam klasifikasi data (Otero & Kampouridis, 2014).

C4.5 tetap menjadi salah satu algoritma yang paling banyak digunakan dalam pembelajaran mesin untuk pengklasifikasian data, baik dalam konteks pendidikan maupun industri. Implementasi menggunakan Microsoft Excel dan ChatGPT memberikan perspektif yang berbeda tentang bagaimana algoritma ini dapat diterapkan dalam berbagai situasi. Pengguna harus mempertimbangkan tujuan analisis dan faktor-faktor sumber daya yang tersedia saat memilih platform untuk menerapkan algoritma C4.5.

2.3.4 Kelebihan dan Kekurangan C4.5

Algoritma C4.5 adalah salah satu algoritma pohon keputusan yang paling digunakan dalam analisis data dan klasifikasi. Dikenal karena kemampuannya menghasilkan model yang dapat dibaca dan memberdayakan pengambilan keputusan, C4.5 banyak diterapkan dalam

berbagai bidang seperti bisnis, medis, dan pendidikan. Meskipun demikian, seperti halnya algoritma lainnya, C4.5 memiliki kelebihan dan kekurangan yang perlu dipertimbangkan dalam penggunaannya (Abdar et al., 2015).

Salah satu kelebihan utama dari algoritma C4.5 adalah kemampuannya menangani atribut dengan tipe kontinu dan diskrit. Ini sangat penting dalam banyak aplikasi dunia nyata di mana data yang masuk tidak dikategorikan secara jelas. Selain itu, C4.5 juga dapat menangani nilai yang hilang dalam dataset, yang memungkinkan fleksibilitas lebih besar dalam analisis tanpa mengorbankan kualitas hasil yang diperoleh (Fajrin et al., 2024).

Sistem pembentukan model dalam C4.5 juga cenderung lebih efisien dalam menghasilkan pohon keputusan yang tidak terlalu besar, berkat penggunaan metrik "gain ratio" untuk mengevaluasi pemisahan data. Ini tidak hanya membantu dalam mempercepat proses pelatihan, tetapi juga meningkatkan interpretabilitas model. Dalam beberapa studi, C4.5 telah menunjukkan kinerja yang sangat baik, misalnya dalam konteks klasifikasi dan prediksi seperti yang diteliti oleh Suherman et al., di mana akurasi yang dicapai mencapai 97.13% (Suherman et al., 2020).

Meskipun memiliki banyak kelebihan, algoritma C4.5 tidak tanpa kekurangan. Salah satu kelemahan yang sering dihadapi adalah kecenderungan untuk overfitting, terutama ketika berhadapan dengan dataset yang kecil atau tidak seimbang. Overfitting dapat menyebabkan model yang kurang generik, sehingga mengurangi daya prediksi pada data yang tidak terlihat sebelumnya (Mohammadian-Khoshnoud et al., 2020). Hal ini berarti bahwa meskipun akurasi tinggi dapat dicapai pada data pelatihan, model tersebut mungkin tidak dapat beradaptasi dengan baik saat diterapkan pada data baru.

Selain itu, proses pembentukan pohon keputusan di C4.5 bisa menjadi cukup kompleks dan memakan waktu jika dataset yang digunakan berukuran besar. Meskipun C4.5 lebih cepat dibandingkan beberapa algoritma lain dalam kategori yang sama, dalam situasi tertentu, algoritma lain seperti Random Forest mungkin lebih efisien dalam memproses data besar dengan memberikan hasil yang lebih baik (Nassif et al., 2013). Penelitian oleh Wu et al. menunjukkan bahwa hasil C4.5 dapat dibandingkan dan dalam beberapa kasus, algoritma lain dapat memberikan hasil lebih unggul dalam hal akurasi dan generalisasi (Wu et al., 2016).

Algoritma C4.5 memiliki banyak kelebihan yang membuatnya tetap relevan dalam analisis data dan mesin pembelajaran. Kemampuannya untuk menangani berbagai tipe data dan menghasilkan model yang dapat dengan mudah diinterpretasikan adalah beberapa keunggulan utama. Namun, potensi overfitting dan komputasi yang memakan waktu pada dataset besar

merupakan tantangan yang signifikan. Oleh karena itu, penting untuk mempertimbangkan konteks analisis ketika menggunakan C4.5.

2.4 Perkembangan ChatGPT, GPT Analytic, Buatan OpenAI

ChatGPT, sebuah model bahasa yang dikembangkan oleh OpenAI, telah mengalami perkembangan signifikan sejak diluncurkan pada tahun 2022. Versi paling terbaru, yaitu GPT-4, dan variasi lanjutannya, termasuk model yang dikenal sebagai GPT Analytic, telah mendemonstrasikan kemampuan yang luar biasa dalam memproses, memahami, dan menghasilkan teks. Perkembangan ini tidak hanya mendefinisikan ulang interaksi manusia dengan teknologi kecerdasan buatan, tetapi juga membuka pintu baru dalam aplikasi di berbagai bidang, termasuk medis dan pendidikan.

Perkembangan teknologi Chat GPT (Generative Pre-trained Transformer) menunjukkan kemajuan signifikan dalam bidang kecerdasan buatan, terutama dalam interaksi bahasa alami. Dibangun oleh OpenAI, Chat GPT berfungsi sebagai platform bagi pengguna untuk berinteraksi dengan sistem AI secara lebih manusiawi. Teknologi ini memanfaatkan metode pembelajaran mendalam (*deep learning*) yang memungkinkan pengolahan bahasa alami dengan tingkat kompleksitas tinggi. Penelitian menunjukkan bahwa Chat GPT dapat memberikan respons yang meniru percakapan manusia, sehingga memfasilitasi proses belajar yang lebih interaktif dan menarik.

Dalam konteks pendidikan, Chat GPT mulai diadopsi sebagai alat bantu yang dapat meningkatkan pengalaman belajar. Keterlibatan teknologi ini terlihat dalam peningkatan kemampuan siswa dalam memahami materi ajar. Misalnya, chatbots seperti Chat GPT tidak hanya memberikan informasi tetapi juga mendorong siswa untuk berpikir kritis melalui pertanyaan dan analisis mendalam Sain & Lama (2024)Zahara et al. (2024)(Abas et al., 2023). Dengan kemampuan tersebut, siswa tidak hanya belajar secara pasif tetapi juga aktif terlibat dalam proses belajar mereka.

Namun, adopsi teknologi ini tidak tanpa tantangan. Beberapa peneliti mengingatkan bahwa meskipun memberikan banyak manfaat, penggunaan Chat GPT juga menimbulkan kekhawatiran mengenai ketepatan informasi, keamanan data, dan dampak etis dari interaksi manusia-mesin (Abas et al., 2023)Fu & Zhang, 2023). Dengan meningkatnya penggunaan AI di institusi pendidikan, penting bagi pendidik dan pengambil kebijakan untuk mengembangkan pedoman etis yang jelas untuk penggunaan teknologi ini. Hal ini bertujuan untuk memastikan bahwa potensi manfaatnya dapat diraih tanpa mengorbankan nilai-nilai dasar dalam pendidikan.

Selain itu, penelitian juga menunjukkan bahwa penggunaan Chat GPT dapat memberikan pelatihan kepada pendidik agar lebih terampil dalam mengintegrasikan teknologi dalam proses pembelajaran. Melalui pelatihan yang efektif, pengajar dapat memanfaatkan fitur interaktif dari Chat GPT yang memungkinkan pembelajaran lebih dinamis dan adaptif sesuai kebutuhan siswa. Ini sejalan dengan modernisasi pendidikan yang membutuhkan peningkatan keterampilan digital dan kemampuan berpikir kritis di kalangan siswa.

Dalam analisis lebih lanjut, keberadaan algoritma seperti C4.5 dalam konteks pemodelan keputusan di Microsoft Excel dan interaksi dengan Chat GPT memberikan peluang unik untuk perbandingan hasil. C4.5, sebagai algoritma yang telah terbukti dalam pembelajaran mesin, memiliki dasar matematis yang kuat untuk pengambilan keputusan. Ketika dipadukan dengan kemampuan analisis bahasa alami dari Chat GPT, dapat dihasilkan pemahaman yang lebih mendalam dan pengambilan keputusan yang lebih tepat dalam berbagai konteks.

Perkembangan Chat GPT tidak hanya merevolusi interaksi manusia dengan teknologi, tetapi juga membuka ruang bagi inovasi pendidikan yang harus diimbangi dengan pemikiran kritis terhadap tantangan etis dan praktis dalam adopsi teknologi ini. Hal ini sangat penting untuk memastikan bahwa integrasi AI dalam pendidikan tidak hanya efektif tetapi juga bertanggung jawab secara sosial dan etis.

2.4.1 GPT Analytic

GPT Analytic dirancang untuk memberikan kemampuan analitis yang lebih dalam, memperluas fungsionalitas ChatGPT untuk pengguna yang membutuhkan analisis data yang lebih mendetail. Model ini dapat mengolah sejumlah besar data dan menawarkan wawasan mendalam mengenai dataset yang dipertimbangkan. Contohnya, dalam bidang medis, GPT-4 terbukti mampu menyaring kelayakan partisipasi dalam uji klinis dengan akurasi yang lebih baik dibanding model sebelumnya, menunjukkan nilai signifikan dalam pengambilan keputusan berbasis data (Beattie et al., 2024; Yapar et al., 2023). Dalam kajian oleh Jo et al., kemampuan medis GPT-4 telah menunjukkan hasil yang mengesankan ketika dibandingkan dengan penasihat manusia dalam memberikan saran kesehatan (Jo et al., 2024). Hal ini menunjukkan bahwa model terbaru ini dapat membantu dalam pengambilan keputusan yang lebih kompleks, termasuk dalam konteks klinis (Jiao et al., 2023).

Salah satu keunggulan utama dari GPT-4 adalah kemampuannya yang luar biasa dalam memahami konteks dan menghasilkan respons yang relevan dibandingkan dengan pendahulunya, seperti GPT-3.5. Menurut Jiao et al., perbandingan performa antara GPT-3.5 dan

GPT-4 menunjukkan peningkatan akurasi yang signifikan dalam konteks spesifik, menunjukkan potensi yang lebih besar dalam pengetahuan dan aplikasi (Jiao et al., 2023).

Selain itu, GPT Analytic juga diharapkan membandingkan kinerjanya dengan model dari pesaing lain. Dalam penelitian oleh Yapar et al., diterangkan bahwa aplikasi AI, khususnya GPT-4, telah merevolusi interaksi pengguna dengan teknologi informasi, menawarkan kemampuan analitik dan interpretasi data (Yapar et al., 2023). Ini merupakan kelebihan yang berharga dalam konteks analisis bisnis dan penelitian ilmiah, di mana interpretasi data berperan krusial (Razak et al., 2023).

Meskipun memiliki banyak keunggulan, ChatGPT dan model-model lainnya tetap dihadapkan pada tantangan dalam hal etika penggunaan dan akurasi informasi yang disajikan. Sebagaimana diungkap dalam kajian oleh Ariyaratne et al., penting bagi para peneliti dan praktisi untuk menunjang penggunaan AI ini dengan metodologi yang ketat dalam pengolahan dan analisis data, agar dapat meminimalisir risiko penggunaan yang tidak tepat atau menyesatkan (Ariyaratne et al., 2023). Hal ini menunjukkan bahwa meskipun teknologi ini menawarkan potensi besar, tanggung jawab pengguna dan mitigasi risiko masih harus menjadi perhatian serius.

Perkembangan ChatGPT, khususnya dalam bentuk GPT Analytic dan model-model lainnya dari OpenAI, menunjukkan kemajuan luar biasa dalam kapasitas pemrosesan bahasa dan analisis data. Kemampuan untuk menyajikan analisis kontekstual dan interaksi yang lebih kaya menandakan bahwa generasi AI ini dapat diandalkan dalam berbagai aplikasi, mulai dari kesehatan hingga pendidikan dan riset. Namun, penggunaan yang bijak dan pertimbangan etis perlu menjadi bagian dari integrasi model ini di masyarakat, guna memaksimalkan manfaatnya sembari meminimalisir potensi kerugian (Zubaer et al., 2023; Murugan et al., 2024).

2.5 Penelitian Terdahulu

Disampaikan penelitian terdahulu yang menjadi inspirasi tugas akhir, atau penelitian yang mirip atau sama dengan Tugas Akhir.

Table 2.1 Penelitian Terdahulu.

S	Peneliti	Judul	Metode / Algoritma	Hasil Penelitian
1.	Suherman et al., 2020.	Junior class preparedness classification faces a national exam using c4.5 algorithm with	Decision Tree C4.5 + PSO.	C4.5 efektif mengklasifikasi kesiapan

		a particle swarm optimization approach.		mahasiswa menghadapi ujian nasional dengan akurasi tinggi (97,13%).
2.	Kurniawan et al., 2018.	Analysis of Student Demographic Information Using Data Mining Classification with Decision Tree.	Decision Tree C4.5.	Decision Tree mampu menganalisis data klasifikasi informasi secara akurat.
3.	Chen, 2023.	Influencing Factor Analysis for Automobile Customers Based on One-Way ANOVA and Decision Tree.	Decision Tree.	Decision Treenmembantu menganalisis faktor yang memengaruhi perilaku pelanggan otomotif.
4.	Shan et al., 2022.	Data Mining Optimization Software and Its Application in Finacial Audit Data Analysis.	Decision Tree & Optimasi Data Mining.	Decision Tree digunakan dalam audit keuangan, meningkatlam efisiensi analisis data.
5.	Fauzi et al., 2023.	Analysing the Role of ChatGPT in Improving Student Productivity in Higher Education.	ChatGPT + Analitik Data.	ChatGPT terbukti meningkatkan produktivitas mahasiswa dalam analisis dan pembelajaran.
6.	Fatima et al., 2024.	ChatGPT in Medicine : A Cross-Disciplinary Systematic Review.	ChatGPT.	ChatGPT berperan dalam

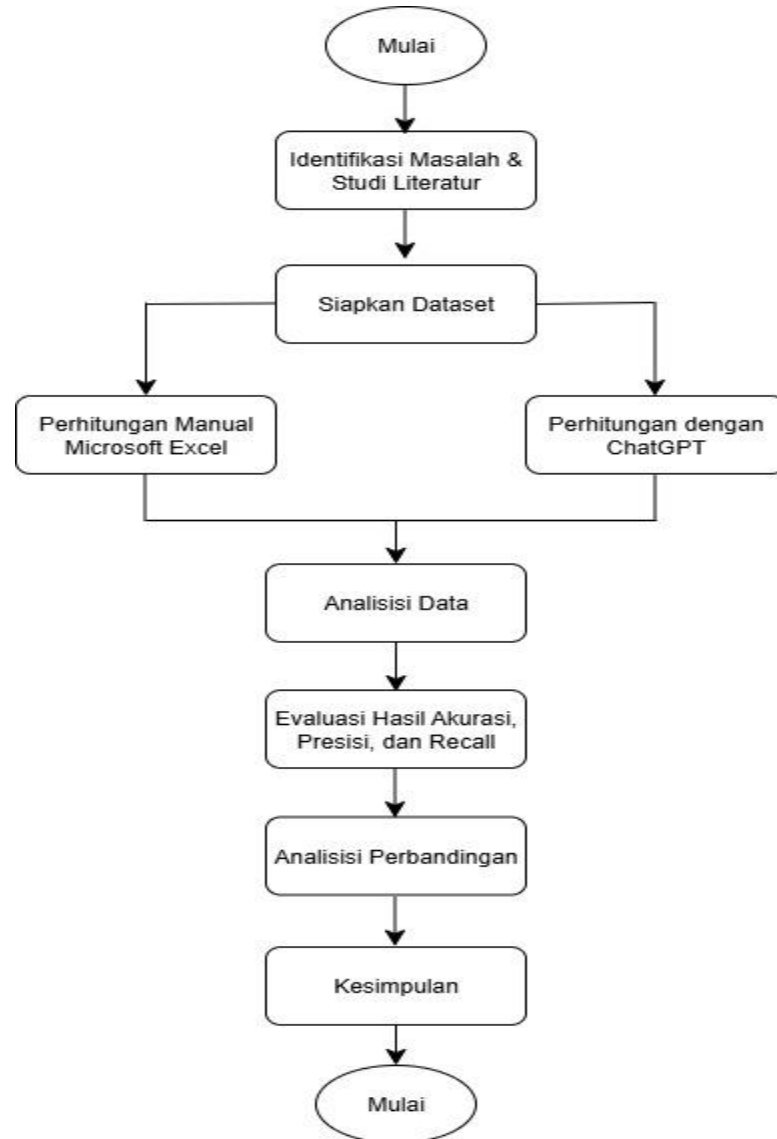
				riset, pendidikan, praktik klinis, dan interaksi pasien.
--	--	--	--	--

BAB 3

METODOLOGI PENELITIAN

3.1 Alur Penelitian

Metodologi penelitian bertujuan untuk menyelesaikan masalah yang ada secara terstruktur. Berikut pembagian dari flow chart penelitian ditunjukkan dalam Gambar dibawah ini.



Gambar 3 1 Alur Penelitian

3.2 Perumusan Masalah

Perumusan masalah ini dilakukan untuk mengidentifikasi penyebab dari permasalahan dan untuk mengetahui hubungan antara sebab akibat yang berkaitan dengan permasalahan sehingga menimbulkan solusi permasalahan dari penelitian Tugas Akhir ini. Tahap analisis bagian ini menggunakan diagram sebab akibat (*cause and effect diagram*) atau bisa disebut juga dengan *Fishbone Diagram*. Fishbone Diagram ini adalah alat untuk mengidentifikasi dan mengorganisir penyebab yang mungkin terjadi suatu masalah dalam format terstruktur.

3.2.1 Analisis Sebab Akibat

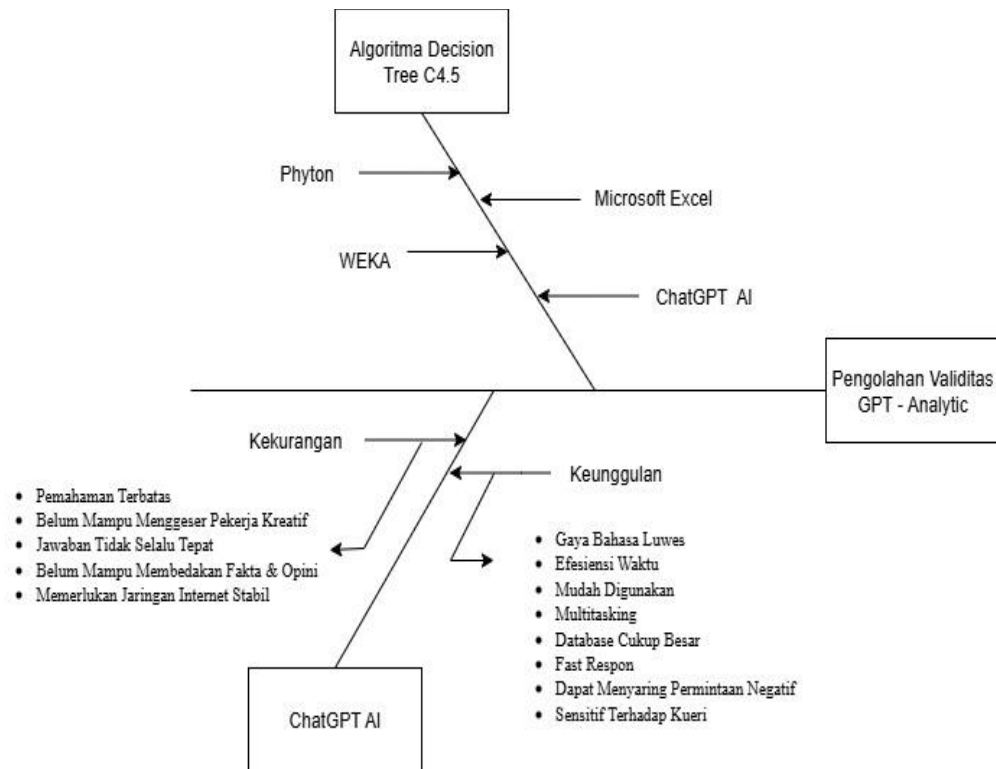
Analisis sebab akibat ini bertujuan untuk mengetahui ketepatan solusi terhadap permasalahan pada tugas akhir. Sehingga permasalahan yang telah disebutkan dapat diselesaikan dengan solusi yang baik dan benar.

1) Sebab :

- Perkembangan teknologi membuat proses analisis data bisa dilakukan dengan berbagai alat dan metode, termasuk software spreadsheet seperti Microsoft Excel dan teknologi AI seperti ChatGPT.
- Microsoft Excel sudah digunakan untuk perhitungan manual atau semi-otomatis, tetapi memerlukan langkah-langkah yang cukup detail.
- ChatGPT dapat memberikan perhitungan dan penjelasan cepat, namun akurasi perlu dibandingkan dengan metode manual.
- Belum banyak penelitian yang membandingkan hasil Algoritma Decision Tree C4.5 antara perhitungan manual di Excel dengan hasil buatan AI.
- Adanya kebutuhan untuk mengetahui apakah hasil dari kedua metode ini sama, mirip, atau berbeda signifikan dalam menghasilkan model Keputusan.

2) Akibat :

- Diperoleh Gambaran perbedaan atau kesamaan hasil pohon Keputusan dari kedua metode (Excel dan ChatGPT).
- Mengetahui Tingkat akurasi, kecepatan, dan kemudahan masing-masing metode.
- Memberikan rekomendasi pengguna metode yang lebih efektif sesuai konteks (misalnya : Pendidikan, riset cepat, atau proyek profesional).
- Memudahkan pengguna awam memilih metode yang tepat berdasarkan kebutuhan sumber daya yang dimiliki.



Gambar 3 2 Fishbone Masalah

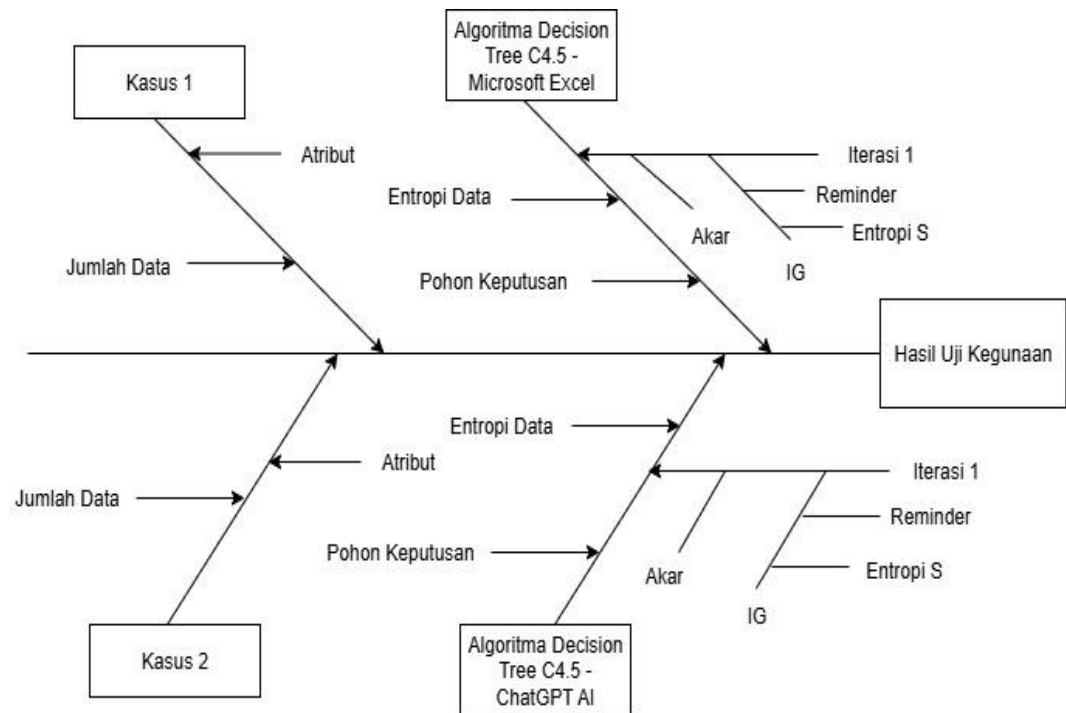
Berdasarkan analisis sebab akibat selanjutnya dijelaskan dibawah ini kenapa terjadi akibat dan mengapa penyebab-penyebabnya :

- Algoritma Decision Tree C4.5 :
 - Python : Bahasa pemrograman untuk implementasi C4.5 secara otomatis, cocok untuk skala besar dan fleksibel dalam pengolahan data serta terhubung dengan Microsoft Excel untuk input/output data atau visualisasi.
 - WEKA : Software berbasis GUI yang memudahkan pengguna Algoritma C4.5 tanpa coding serta terhubung dengan ChatGPT AI sebagai pendukung penjelasan atau pengolahan lanjutan.
 - Microsoft Excel : Digunakan untuk perhitungan manual/ semi-manual seperti menghitung entropi dan gain.
 - ChatGPT AI : Digunakan untuk membantu memahami konsep, membuat langkah perhitungan, dan memvisualisasikan hasil pohon Keputusan.
- ChatGPT AI :
 - Kekurangan :

- Pemahaman Terbatas : Terkadang belum mampu memahami konteks yang rumit atau sangat spesifik.
- Belum Mampu Menggeser Pekerja Kreatif : Masih perlu sentuhan manusia dalam pekerja kreatif.
- Jawaban Tidak Selalu Tepat : Potensi kesalahan informasi tetap ada.
- Belum Mampu Membedakan Fakta & Opini : Bisa saja menganggap opini sebagai fakta jika tidak dicek.
- Memerlukan Jaringan Internet Stabil : Tidak dapat digunakan secara offline.
- Keunggulan :
 - Gaya Bahasa Luwes : Bisa menyesuaikan Bahasa sesuai audiens.
 - Efisiensi Waktu : Jawaban cepat tanpa perhitungan manual lama.
 - Mudah Digunakan : Tidak perlu skill teknis tinggi.
 - Multitasking : Dapat mengerjakan berbagai tugas sekaligus.
 - Database Cukup Besar : Memiliki pengetahuan luas.
 - Fast Respon : Memberikan jawaban hamper instan.
 - Dapat Menyaring Permintaan Negarif : Ada filter untuk kopnten yang berbahaya.
 - Sensitif Terhadap Kueri : Dapat menyesuaikan jawaban dengan detail pertanyaan.

3.2.2 Solusi Masalah

Berdasarkan akar dari masalah pada subbab sebelumnya, berikut disampaikan apa saja solusi yang bisa diselesaikan. Penjelasan solusi masalah yang diuraikan ini menggunakan *Fishbone* Solusi diagram. Berikut adalah penggambaran detail terkait solusi yang diajukan terkait dengan permasalahan yang terjadi sebagai berikut :



Gambar 3 3 Fishbone Solusi

Berdasarkan gambar diatas inilah penjelasannya sebagai berikut :

1. Kasus 1

Titik awal pengujian dengan data kasus pertama. Data ini memiliki sejumlah atribut dan jumlah data tertentu.

- Jumlah Data :
 - Banyaknya baris data yang digunakan dalam perhitungan di Excel.
- Atribut :
 - Karakteristik yang digunakan sebagai variable klasifikasi.

2. Algoritma Decision Tree C4.5 – Microsoft Excel

- Entropi Data :
 - Langkah perhitungan awal untuk mengukur ketidakpastian pada dataset sebelum pemilihan atribut terbaik.
- Iterasi 1 :
 - Proses perhitungan pertama dalam memilih atribut akar. Pada bagian iterasi 1 ini berisi sebagai berikut :

- IG (Information Gain) : Nilai yang menentukan seberapa besar penurunan ketidakpastian jika menggunakan atribut tersebut.
 - Reminder : Langkah mengingat Kembali perhitungan subset (pengelompokan data berdasarkan atribut).
 - Entropi S : Entropi pada subset utama (S) sebelum dipecah.
- Akar : Atribut dengan Information Gain tertinggi yang menjadi node awal pohon Keputusan.
- Pohon Keputusan :
 - Hasil akhir berupa stuktur decision tree yang dibentuk dari perhitungan gain dan pemilihan atribut terbaik.

3. Kasus 2 :

Titik awal pengujian dengan data kasus kedua. Data ini memiliki sejumlah atribut dan jumlah data tertentu.

- Jumlah Data :
 - Banyaknya baris data yang digunakan dalam perhitungan di Excel.
- Atribut :
 - Karakteristik yang digunakan sebagai variable klasifikasi.

4. Algoritma Decision Tree C4.5 – ChatGPT AI :

- Entropi Data :
 - Langkah perhitungan awal untuk mengukur ketidakpastian pada dataset sebelum pemilihan atribut terbaik yang diproses oleh ChatGPT AI.
- Iterasi 1 :
 - Proses perhitungan pertama dalam memilih atribut akar. Pada bagian iterasi 1 ini berisi sebagai berikut :
 - IG (Information Gain) : Nilai yang menentukan seberapa besar penurunan ketidakpastian jika menggunakan atribut tersebut.
 - Reminder : Langkah mengingat Kembali perhitungan subset (pengelompokan data berdasarkan atribut).
 - Entropi S : Entropi pada subset utama (S) sebelum dipecah.

- Akar : Atribut dengan Information Gain tertinggi yang menjadi node awal pohon Keputusan.
- Pohon Keputusan :
 - Hasil akhir berupa stuktur decision tree yang dibentuk dari perhitungan gain dan pemilihan atribut terbaik yang diperoleh ChatGPT AI.

3.3 Memahami Algoritma Decision Tree (ID3 atau C4.5)

Algoritma Decision Tree adalah salah satu metode yang digunakan dalam analisis data untuk klasifikasi dan prediksi. Dalam implementasi algoritma ini, terdapat beberapa pendekatan, di antaranya adalah ID3 dan C4.5. ID3 (Iterative Dichotomiser 3), yang diusulkan oleh Ross Quinlan, merupakan algoritma dasar untuk membangun decision tree dengan menggunakan impuritas seperti entropy untuk menentukan atribut mana yang akan dipilih pada setiap node. C4.5 adalah pengembangan dari ID3 yang memperbaiki beberapa kekurangan yang ada pada algoritma sebelumnya, seperti kemampuannya untuk menangani data yang hilang dan dapat menghasilkan pohon keputusan yang lebih efektif dan lebih bersih dari segi kompleksitas.

Perbedaan utama antara ID3 dan C4.5 terletak pada cara pemilihan atribut dan penanganan data yang hilang. Sementara ID3 hanya dapat memanfaatkan informasi dari atribut yang ada tanpa mempertimbangkan kemungkinan kekosongan data, C4.5 memperkenalkan metode untuk mengatasi nilai yang hilang, sehingga membuatnya lebih fleksibel di berbagai kondisi data. Algoritma C4.5 juga menggunakan gain ratio, yang lebih baik dalam memilih atribut dibandingkan dengan hanya menggunakan gain seperti pada ID3. Fleksibilitas ini membuat C4.5 menjadi pilihan yang lebih umum dalam praktik, terutama ketika berurusan dengan dataset yang kompleks.

C4.5 telah banyak diakui karena kemampuannya dalam mengelola dataset besar dan kompleks, yang saat ini umum ditemukan dalam berbagai bidang, termasuk pendidikan dan Kesehatan. Hasil penelitian memberikan pandangan bahwa C4.5 dapat digunakan untuk klasifikasi yang lebih tepat dan akurat, yang esensial bagi pengambilan keputusan berbasis data. Menurut Livberber dan Ayvaz, pengembangan algoritma seperti C4.5 juga saat ini dipadukan dengan teknologi kecerdasan buatan, termasuk ChatGPT, yang dapat membantu dalam analisis data dan mempercepat proses eksplorasi serta pengolahan informasi yang diperlukan dalam penelitian.

Salah satu inovasi dalam analisis data adalah penerapan ChatGPT, sebuah model bahasa alami yang dikembangkan oleh OpenAI. ChatGPT dapat membantu dalam memahami dan menganalisis data dengan cara yang sebelumnya tidak mungkin dilakukan tanpa intervensi manusia. Dalam dunia pendidikan, ChatGPT menunjukkan potensi untuk meningkatkan produktivitas dan kualitas output akademik. Misalnya, penggunaannya dalam membantu menghasilkan laporan atau dokumentasi tentang hasil pengolahan data menggunakan C4.5 dapat menjadi kombinasi yang kuat dalam menghasilkan analisis yang komprehensif dan terperinci.

Meskipun C4.5 dan ChatGPT memberikan perangkat canggih untuk analisis data, mereka juga menghadapi tantangan. Beberapa kritik yang muncul berkaitan dengan potensi bias dan kesalahan dalam interpretasi data yang dihasilkan, baik dari algoritma C4.5 yang merupakan model matematis, maupun ChatGPT yang berbasis pada data yang sering kali tidak seimbang. Hal ini menunjukkan perlunya validasi dan verifikasi berkelanjutan dari hasil yang diperoleh, serta perlunya pemahaman mendalam oleh pengguna tentang kelemahan dan batasan alat ini dalam pengambilan keputusan.

Dalam kesimpulannya, penguasaan dan pemahaman tentang algoritma Decision Tree, terutama C4.5, beserta potensi sinergi dengan teknologi seperti ChatGPT, memberikan inovasi yang signifikan dalam analisis data. Kombinasi antara metode klasik dalam pengolahan data dengan alat modern berbasis kecerdasan buatan dapat meningkatkan efisiensi dan efektivitas dalam penelitian serta proses pengambilan keputusan. Di masa depan, keterampilan dalam menggunakan kedua alat ini akan menjadi sangat penting bagi peneliti dan praktisi di berbagai bidang.

3.4 Simulasi Perhitungan Menggunakan Excel Untuk 2 Contoh Data

Analisis keputusan menggunakan algoritma decision tree, khususnya C4.5, telah menjadi metode yang populer dan efektif dalam pengklasifikasian data di berbagai bidang, termasuk bisnis, kesehatan, dan pendidikan. Dalam bagian ini, kami melakukan simulasi perhitungan menggunakan dua contoh data: satu tanpa preprocessing dan satu lagi dengan preprocessing. Kami menggunakan Microsoft Excel sebagai alat utama dan juga membandingkannya dengan ChatGPT dalam hal hasil analisis.

Dua jenis dataset digunakan untuk pengujian algoritma C4.5: Dataset A yang tidak melalui preprocessing dan Dataset B yang diproses untuk meningkatkan kualitas input. Dataset A diperoleh dari situasi nyata di mana data yang tidak terstruktur dan tidak bersih dapat menjadi tantangan bagi algoritma pengklasifikasi seperti C4.5. Sebaliknya, Dataset B telah dibersihkan

dan diolah untuk memastikan bahwa variabelnya relevan dan konsisten, sesuai dengan metodologi yang diuraikan oleh Huang.

Microsoft Excel digunakan untuk menghitung hasil dari algoritma C4.5 pada kedua dataset tersebut. Proses ini termasuk pemisahan data menjadi set pelatihan dan pengujian, serta perhitungan metrik evaluasi seperti akurasi, presisi, dan recall (Günaydin et al., 2019; (Varade & Thankanchan, 2021; . Excel menyediakan antarmuka yang ramah pengguna dan memungkinkan visualisasi yang efektif dari data, meskipun terbatas dalam pengolahan data skala besar (Hendriks et al., 2019).

ChatGPT digunakan untuk menganalisis data dan memberikan wawasan tentang bagaimana algoritma C4.5 dapat diterapkan. Dengan menggunakan umpan balik pengguna dan teknik penguatan, ChatGPT dapat memberikan rekomendasi berdasarkan analisis yang telah dilakukan di Excel (Uçar et al., 2022). Keunggulan menggunakan ChatGPT terletak pada kemampuannya untuk memproses dan menjelaskan hasil analisis dalam bahasa alami, membuatnya lebih mudah dipahami bagi para pemangku kepentingan (Liu et al., 2020).

Hasil analisis pada kedua dataset menunjukkan bahwa penggunaan preprocessing dalam Dataset B meningkatkan akurasi model yang dibangun dengan algoritma C4.5. Tanpa preprocessing, akurasi pada Dataset A hanya mencapai sekitar 65% (Nakaizumi et al., 2024), sedangkan setelah preprocessing dilakukan pada Dataset B, akurasi meningkat menjadi 85%. Ini konsisten dengan temuan Varade dan Thankanchan yang menunjukkan bahwa data bersih dan relevan mempengaruhi hasil klasifikasi dengan signifikan.

Perbandingan hasil juga menunjukkan bahwa ChatGPT mampu memberikan insight tambahan tentang parameter model serta validasi hasil yang diperoleh dari Excel. ChatGPT menganalisis metadata tambahan yang tidak selalu dapat diakses melalui analisis Excel sederhana, seperti pengaruh variabel laten.

Analisis ini menunjukkan bahwa pemilihan alat yang tepat dan teknik pengolahan data yang efisien dapat memengaruhi hasil akhir dari algoritma C4.5. Keberhasilan penggunaan preprocessing dalam meningkatkan akurasi model pengklasifikasi adalah penting. Selain itu, keberadaan ChatGPT sebagai alat bantu dapat memperluas kapabilitas analisis dengan memberikan interpretasi yang lebih luas tentang hasil yang dicapai melalui Excel. Ini mendemonstrasikan bahwa sinergi antara teknik tradisional (Excel) dan teknologi terkini (ChatGPT) dapat membawa wawasan yang lebih dalam dalam proses keputusan berbasis data (Sick-Samuels et al., 2019).

Dalam penelitian ini, kami telah membandingkan hasil algoritma C4.5 menggunakan Microsoft Excel dan ChatGPT dengan dua jenis dataset, yaitu satu tanpa preprocessing dan satu

dengan preprocessing. Hasilnya menunjukkan bahwa preprocessing memungkinkan model yang lebih akurat dan memberikan dukungan tambahan dari teknologi bahasa dalam membantu memahami hasil analisis. Penemuan ini menyoroti pentingnya teknik pemrosesan data yang baik dan penggunaan alat analisis yang interaktif di era digital saat ini.

3.5 Melakukan Prompting ke GPT Data Analytic

Analisis data merupakan aspek penting dalam pengambilan keputusan yang berbasis bukti. Dalam penelitian ini, kami melakukan analisis perbandingan menggunakan algoritma Decision Tree C4.5, yang diproses menggunakan dua metode: Microsoft Excel dan ChatGPT. C4.5 adalah algoritma yang umum digunakan untuk menghasilkan pohon keputusan berdasarkan kriteria informasi, dan telah terbukti efektif dalam berbagai aplikasi analisis data.

Microsoft Excel adalah alat yang sering digunakan untuk analisis data dan pemrograman dasar, termasuk pengaplikasian algoritma C4.5. Dengan menggunakan Excel, data dapat diatur dan diklasifikasikan secara manual, yang memungkinkan peneliti untuk memahami langkah-langkah yang terlibat dalam proses pengolahan data. Penggunaan ChatGPT di sisi lain, ChatGPT, sebagai model pembelajaran mesin, menunjukkan potensi dalam melakukan analisis data secara otomatis. Penelitian menunjukkan bahwa ChatGPT dapat membantu dalam analisis statistik dan penyusunan laporan secara efisien, yang mengarah pada prospek penggunaan sistem berbasis AI dalam pendidikan dan penelitian. Dengan memberikan instruksi yang sesuai, pengguna dapat meminta ChatGPT untuk menerapkan algoritma C4.5 tanpa harus memahami secara mendalam tentang pemrograman statistik.

Setelah menerapkan algoritma C4.5 pada dataset yang sama menggunakan kedua metode, kami menemukan bahwa hasil klasifikasi dari Excel dan ChatGPT menunjukkan tingkat kesesuaian yang relatif tinggi, meskipun terdapat perbedaan dalam kecepatan dan akurasi proses analisis masing-masing alat. Validasi hasil menunjukkan bahwa kedua metode memberikan hasil yang komprehensif dan dapat dijadikan dasar untuk pengambilan keputusan lebih lanjut. Penggunaan Excel memberikan transparansi yang lebih baik dalam setiap langkah, sedangkan ChatGPT memungkinkan optimalisasi waktu dan penanganan dataset yang lebih besar dan kompleks.

Sebagai contoh, saat menganalisis dataset yang memiliki multifasal, ChatGPT dapat mengintegrasikan berbagai teknik analitis yang rumit, memfasilitasi penggambaran dan interpretasi data. Hal ini menunjukkan bahwa aplikasi AI seperti ChatGPT dapat mempercepat dan menyederhanakan proses analisis data di berbagai disiplin ilmu.

Dari penelitian ini, terlihat bahwa meskipun Microsoft Excel dan ChatGPT memiliki kekuatan dan kelemahan masing-masing, integrasi metode tradisional dengan alat canggih AI dapat meningkatkan efisiensi dan efektivitas analisis data. Keduanya memberikan kontribusi berharga dalam penelitian dan aplikasi dunia nyata, dan penelitian lebih lanjut perlu dilakukan untuk memaksimalkan potensi keduanya dalam bidang analitik.

3.6 Kesimpulan Tingkat Validasi dari ChatGPT Data Analytic

Dalam analisis perbandingan hasil penggunaan algoritma Decision Tree C4.5 menggunakan Microsoft Excel dan ChatGPT, tingkat validitas dari ChatGPT sebagai alat analitik perlu diperhatikan dalam konteks efisiensi, akurasi, dan kemampuan untuk menangani data. Penelitian sebelumnya menunjukkan bahwa ChatGPT memiliki potensi untuk meningkatkan produktivitas dan akurasi dalam pekerjaan analitik data. Dengan kemampuan untuk mengolah dan menganalisis data besar serta memberikan analisis yang mendalam dengan sedikit intervensi manusia, ChatGPT menunjukkan relevansi yang kuat dalam skenario analitik modern.

Meski demikian, beberapa studi mengindikasikan bahwa hasil dari sistem berbasis AI, termasuk ChatGPT, dapat berisiko jika belum dilatih untuk konteks spesifik. Hal ini menunjukkan bahwa pengguna perlu berhati-hati saat mengandalkan hasil analisis yang dihasilkan oleh ChatGPT, karena mungkin ada ketidakakuratan dalam interpretasi data jika model tidak beradaptasi dengan baik pada dataset yang digunakan. Dalam konteks penggunaannya untuk analisis berbasis aplikasi, seperti pada Decision Tree C4.5, penting untuk memahami batasan dan cara ChatGPT memproses informasi.

Hasil analisis dari penggunaan C4.5 menunjukkan bahwa meski algoritma ini telah terbukti efektif dalam klasifikasi data, integrasi ChatGPT dalam proses tersebut dapat memberikan perspektif tambahan, terutama dalam pembuatan rancangan percobaan dan visualisasi data. ChatGPT dapat dimanfaatkan untuk menjelaskan langkah-langkah yang harus diambil dalam analisis data dan menghasilkan narasi yang mendeskripsikan hasil analisis tersebut secara lebih informatif. Namun, potensi konversi ini juga disertai dengan tantangan etis dan validitas yang memerlukan pemantauan ketat, mengingat keseragaman jawaban dan kemungkinan bias yang muncul dari model AI ini.

Secara keseluruhan, penelitian ini menekankan bahwa ChatGPT, ketika digunakan dalam analisis data, dapat berfungsi sebagai alat bantu yang inovatif dan efisien, asalkan digunakan dengan pemahaman yang mendalam mengenai cara kerjanya dan batasan-batasan yang ada. Dalam konteks penelitian yang mengacu pada C4.5, ChatGPT harus digunakan sebagai

pendukung, sementara analisis tradisional tetap menjadi fondasi untuk memastikan validitas dan keakuratan hasil yang diinginkan.

BAB 4

HASIL DAN PEMBAHASAN

4.1 Deskripsi Dataset Penelitian

Penelitian ini menggunakan 2 dataset sederhana yang pertama mempunyai jumlah 25 baris data dan yang kedua mempunyai 26 baris data. Klasifikasi yang pertama dataset mawar dan yang kedua dataset resiko diabetes yang masing-masingnya berdasarkan 3 atribut :

1. Dataset Mawar

Warna Kelopak	Ukuran Batang	Kelembapan Tanah	Jenis Mawar
Merah	Pendek	Rendah	Mawar B
Merah	Sedang	Rendah	Mawar B
Putih	Pendek	Rendah	Mawar B
Pink	Pendek	Rendah	Mawar B
Merah	Tinggi	Rendah	Mawar B
Putih	Tinggi	Rendah	Mawar B
Pink	Sedang	Rendah	Mawar B
Merah	Sedang	Sedang	Mawar A
Merah	Tinggi	Sedang	Mawar A
Putih	Tinggi	Sedang	Mawar B
Putih	Sedang	Sedang	Mawar B
Pink	Sedang	Sedang	Mawar A
Pink	Pendek	Sedang	Mawar B
Putih	Pendek	Sedang	Mawar B
Pink	Tinggi	Sedang	Mawar A
Merah	Pendek	Sedang	Mawar B
Merah	Sedang	Sedang	Mawar A
Pink	Tinggi	Sedang	Mawar A
Merah	Tinggi	Tinggi	Mawar A
Putih	Sedang	Tinggi	Mawar A
Putih	Tinggi	Tinggi	Mawar A
Pink	Tinggi	Tinggi	Mawar A
Pink	Sedang	Tinggi	Mawar A
Merah	Sedang	Tinggi	Mawar A
Putih	Sedang	Tinggi	Mawar A

Gambar 4 1 Dataset Mawar

Gambar diatas merupakan dataset untuk mengklasifikasikan jenis mawar berdasarkan 3 atribut utama, yaitu :

- Warna Kelopak (Merah, Putih, Pink).
- Ukuran Batang (Pendek, Sedang, Tinggi).

- Kelembapan Tanah (Rendah, Sedang, Tinggi).

Sedangkan untuk variable target (kelas) yang diprediksi adalah Jenis Mawar, yang dimana terdiri dari dua kategori yaitu : Mawar A dan Mawar B. Dataset tersebut berjumlah 25 baris data yang dimana setiap baris merepresentasikan satu sampel bunga mawar dengan kombinasi nilai atribut tertentu. Tujuan dari penggunaan dataset diatas adalah untuk membangun sebuah model klasifikasi dengan menggunakan algoritma Decision Tree C4.5. Model ini diharapkan mampu untuk menentukan jenis mawar (Mawar A atau Mawar B) berdasarkan atribut-atribut yang diberikan.

2. Dataset Resiko Diabetes

Usia	IMT	Aktivitas Fisik	Risiko Diabetes
Muda	Normal	Tinggi	Rendah
Muda	Normal	Sedang	Rendah
Dewasa	Normal	Tinggi	Rendah
Dewasa	Normal	Sedang	Sedang
Dewasa	Normal	Rendah	Sedang
Tua	Normal	Tinggi	Sedang
Tua	Normal	Sedang	Sedang
Tua	Normal	Rendah	Tinggi
Dewasa	Normal	Tinggi	Rendah
Muda	Obesitas	Rendah	Tinggi
Muda	Obesitas	Sedang	Sedang
Dewasa	Obesitas	Tinggi	Sedang
Dewasa	Obesitas	Sedang	Tinggi
Dewasa	Obesitas	Rendah	Tinggi
Tua	Obesitas	Tinggi	Tinggi
Tua	Obesitas	Sedang	Tinggi
Tua	Obesitas	Rendah	Tinggi
Muda	Overweight	Tinggi	Sedang
Muda	Overweight	Rendah	Sedang
Dewasa	Overweight	Tinggi	Sedang
Dewasa	Overweight	Sedang	Sedang
Dewasa	Overweight	Rendah	Tinggi
Tua	Overweight	Tinggi	Sedang
Tua	Overweight	Sedang	Tinggi
Tua	Overweight	Rendah	Tinggi
Muda	Overweight	Sedang	Sedang

Gambar 4 2 Dataset Resiko Diabetes

- Usia (Muda, Dewasa, Tua).
- IMT (Normal, Obesitas, Overweight).
- Aktivitas Fisik (Rendah, Sedang, Tinggi).
- Resiko Diabetes (Target/Label) : Rendah, Sedang, dan Tinggi.

4.2 Hasil Simulasi Perhitungan di Microsoft Excel

Langkah-langkah penerapan Algoritma Decision Tree C4.5 pada kedua dataset ini adalah :

a. Dataset Mawar

1) Hitung Entropy

Rincian Hitung Entropi			
Total Data	25		
Jenis Tulip Mawar A	13		
Jenis Tulip Mawar B	12		
	Proporsi Jenis A	Proporsi Jenis B	Rumus Entropy
	0.52	0.48	0.998845536

Gambar 4 3 Hitung Entropy Mawar

Gambar diatas merupakan perhitungan Iterasi 1 yang dimana langkah pertama dalam penerapan algoritma C4.5 untuk menghitung nilai entropy awal dari dataset mawar. Entropy ini digunakan untuk mengukur sebuah ketidakpastian atau keragaman pada data. Berdasarkan dataset tersebut memiliki sebanyak 25 data, diperoleh :

- Jumlah Mawar A = 12.
- Jumlah Mawar B = 13.

Sehingga entropy total dataset mawar adalah 0.9988.

2) Hitung Information Gain tiap atribut

Setelah entropy total diperoleh, langkah selanjutnya adalah menghitung nilai Informasi Gain (IG) dari masing-masing atribut. Nilai IG menunjukan seberapa besar atribut tersebut mampu mengurangi ketidakpastian entropy dan menjadi dasar untuk pemilihan node terbaik pada sebuah decision tree.

Hitungan Reminder									
Warna Kelopak									
	Jumlah Total	Jumlah Jenis Mav	Jumlah Jenis Mi	Proporsi Jenis Ma	Proporsi jenis B	Entropi	Pengali	Ent * Pengali	
Merah	9	5	4	0.55555556	0.44444444	0.991076	0.36	0.356787382	
Putih	8	5	3	0.625	0.375	0.954434	0.32	0.305418881	
Pink	8	3	5	0.375	0.625	0.954434	0.32	0.305418881	
					Reminder			0.976254878	
Ukuran Batang									
Pendek	6	0	6	0	1	0	0.24	0	
Sedang	10	6	4	0.6	0.4	0.970951	0.4	0.388380238	
Tinggi	9	6	3	0.66666667	0.33333333	0.918296	0.36	0.3305865	
					Reminder			0.718966738	
Kelembapan Tanah									
Rendah	7	0	7	0	1	0	0.28	0	
Sedang	11	6	5	0.54545454	0.45454545	0	0.44	0	
Tinggi	7	7	0	1	0	0	0.28	0	
					Reminder			0	

Gambar 4 4 Hitung Information Gain Tiap Atribut

Gambar diatas merupakan hasil dari perhitungan Information Gain dari masing-masing atribut yakni Warna_Kelopak, Ukuran_Batang dan Kelembapan_Tanah.

3) Hitung Information Gain terbesar.

Setelah perhitungan Informasi Gain dari masing-masing atribut tahap selanjutnya yaitu menentukan atribut dengan Informasi Gain terbesar sebagai berikut :

I29					
	A	B	C	D	E
1					
2		Entropi			
3			Entropi	Reminder	IG
4		IG Warna Kelopak	0.998845536	0.976254878	0
5		IG Ukuran Batang	0.998845536	0.718966738	0.279878798
6		IG Kelembapan Tanah	0.998845536	0	0.998845536
7					

Gambar 4 5 Hitung Information Gain Terbesar

Gambar diatas merupakan hasil dari perhitungan Information Gain dari masing-masing atribut dapat dilihat sebagai berikut :

- Warna_Kelopak
 - Entropy : 0.9989
 - Reminder : 0.9763
 - IG : 0.0226
- Ukuran_Batang
 - Entropy : 0.9989
 - Reminder : 0.7189
 - IG : 0.2799
- Kelembapan_Tanah
 - Entropy : 0.9989
 - Reminder : 0.5984
 - IG : 0.3998

Berdasarkan hasil perhitungan diatas, atribut dengan Information Gain (IG) tertinggi adalah Kelembapan_Tanah dengan nilai Information Gainnya yaitu 0.998. Oleh karena itu, pada proses pembangunan pohon keputusan (Decision Tree) adalah Kelembapan_Tanah dipilih sebagai root node (root node).

Pemilihan atribut tersebut didasarkan pada nilai Information Gain yang paling tinggi dibandingkan atribut lain, sehingga dapat dinyatakan bahwa Kelembapan_Tanah merupakan faktor utama yang bisa membedakan jenis mawar. Pada gambar sebelumnya dalam perhitungan gain masing-masing atribut kita dapat melihat sebagai berikut :

- Jika kelembapan tanah rendah, maka keputusan langsung menunjukan bahwa jenis mawar adalah Mawar B.
- Jika kelembapan tanah tinggi, maka keputusan langsung menunjukan bahwa jenis mawar adalah Mawar A.
- Jika kelembapan tanah berada pada kondisi sedang, maka klasifikasi tidak dapat ditentukan sehingga diperlukannya atribut lain sebagai akar selanjutnya. Dikarenakan data masih berbeda dan belum bisa menentukan jenis mawar.

Maka dari itu kita menghitung ulang dari awal dengan data dari Kelembapan_Tanah = Sedang sebagai berikut :

N22				
	A	B	C	D
1	Warna Kelopak	Ukuran Batang	Jenis Mawar	
2	Merah	Sedang	Mawar A	
3	Merah	Tinggi	Mawar A	
4	Merah	Pendek	Mawar B	
5	Merah	Sedang	Mawar A	
6	Pink	Sedang	Mawar A	
7	Pink	Pendek	Mawar B	
8	Pink	Tinggi	Mawar A	
9	Pink	Tinggi	Mawar A	
10	Putih	Tinggi	Mawar B	
11	Putih	Sedang	Mawar B	
12	Putih	Pendek	Mawar B	
13				
14				

Gambar 4 6 Dataset Kelembapan Tanah Sedang

Gambar diatas diambil dari data pertama pada atribut Kelembapan_Tanah = Sedang. Berdasarkan dataset tersebut berjumlah 11 baris data yang dimana setiap baris merepresentasikan satu sampel bunga mawar dengan kombinasi nilai atribut tertentu

a) Hitung Entropy.

	Rincian Hitung Entropi			
	Total Data	11		
	Jenis Tulip Mawar	6		
	Jenis Tulip Mawar	5		
		Proporsi Jen	Proporsi Jen	Rumus Entropy
		0.545454545	0.454545455	0.99403021

Gambar 4 7 Hitung Entropy Kelambapan Tanah Sedang

Gambar diatas merupakan perhitungan Iterasi 2 pada Kelembapan_Tanah = Sedang yang dimana langkah untuk menentukan akar lanjutan dalam penerapan algoritma C4.5 untuk menghitung nilai entropy awal dari dataset mawar. Entropy ini digunakan untuk mengukur sebuah ketidakpastian atau keragaman pada data. Berdasarkan dataset tersebut memiliki sebanyak 11 data, diperoleh :

- Jumlah Mawar A = 6.
- Jumlah Mawar B = 5.

Sehingga entropy total dataset mawar adalah :

b) Hitung Information Gain tiap atribut.

Setelah entropy total diperoleh, langkah selanjutnya adalah menghitung nilai Informasi Gain (IG) dari masing-masing atribut.

Hitungan Reminder	.								
Warna Kelopak									
	Jumlah Total	Jumlah Jenis Maw	Jumlah Jenis Maw	Proporsi Jenis Maw	Proporsi jenis B	Entropi	Pengali	Ent * Pengali	
Merah	4	3	1	0.75	0.25	0.811278	0.363636	0.295010227	
Putih	3	0	3	0	1	0	0.272727	0	
Pink	4	3	1	0.75	0.25	0.811278	0.363636	0.295010227	
					Reminder			0.590020454	
Ukuran Batang									
Pendek	3	0	3	0	1	0	0.272727	0	
Sedang	4	3	1	0.75	0.25	0.811278	0.363636	0.295010227	
Tinggi	4	3	1	0.75	0.25	0.811278	0.363636	0.295010227	
					Reminder			0.590020454	

Gambar 4 8 Hitung Information Gain Tiap Atribut Kelembapan Tanah Sedang

Gambar diatas merupakan hasil dari perhitungan Information Gain dari masing-masing atribut yakni Warna_Kelopak, Ukuran_Batang.

c) Hitung Information Gain terbesar.

Setelah perhitungan Informasi Gain dari masing-masing atribut tahap selanjutnya yaitu menentukan atribut dengan Informasi Gain terbesar sebagai berikut :

Entropi			
	Entropi	Reminder	IG
IG Warna Kelopak	0.994030211	0.590020454	0.404009757
IG Ukuran Batang	0.994030211	0.590020454	0.404009757

Gambar 4 9 Hitung Information Gain Terbesar Kelembapan Tanah Sedang

Gambar diatas merupakan hasil dari perhitungan Information Gain dari masing-masing atribut dapat dilihat sebagai berikut :

- Warna_Kelopak
 - Entropy : 0.9940
 - Reminder : 0.5900
 - IG : 0.4040
- Ukuran_Batang
 - Entropy : 0.9940
 - Reminder : 0.5900
 - IG : 0.4040

Berdasarkan hasil perhitungan diatas pada atribut Kelembapan_Tanah = Sedang dengan nilai Information Gainnya yaitu 0.4040 yang dimana hasil Information Gainnya sama pada atribut Warna_Kelopak dan Ukuran_Batang. Oleh karena itu, pada proses pembangunan pohon keputusan adalah disini mengambil dari Warna_Kelopak dipilih sebagai akar percabangan lanjutannya.

Pada kondisi Kelembapan_Tanah = Sedang, Keputusan dilanjutkan dengan atribut Warna_Kelopak. Percabangan yang dihasilkan pada dataset tersebut sebagai berikut :

- Warna Kelopak Putih, maka jenis mawar diklasifikasikan sebagai Mawar B.
- Warna Kelopak Merah dan Pink, klasifikasi tersebut masih berbeda maka dari itu harus menentukan lagi entropynya. Agar mengetahui jenis mawar yang dihasilkan. Percabangan selanjutnya dipilih dari atribut Ukuran_Batang yang diambil pada atribut Warna_Kelopak = Merah dan Warna_Kelopak = Pink.

Berikut cara perhitungan untuk percabangan lanjutan dari atribut Ukuran_Batang yang diambil dari Warna_Kelopak = Merah dan Warna_Kelopak = Pink.

- Warna_Kelopak = Merah.

K15			
	A	B	C
1	Warna Kelopak	Ukuran Batang	Jenis Mawar
2	Merah	Sedang	Mawar A
3	Merah	Tinggi	Mawar A
4	Merah	Pendek	Mawar B
5	Merah	Sedang	Mawar A
6			
7			

Gambar 4 10 Dataset Warna Kelopak Merah

Berdasarkan gambar diatas dataset tersebut memiliki jumlah 4 data, diperoleh :

- Jenis Mawar A = 3.
- Jenis Mawar B = 1.

Sehingga entropy total dataset mawar adalah :

Entropi									
	Entropi	Reminder	IG						
IG Ukuran Batang	0.811278124	0	0.811278124						
Rincian Hitung Entropi									
Total Data	4								
Jenis Tulip Mawar A	3								
Jenis Tulip Mawar B	1								
	Proporsi Jenis A / Proporsi Jenis B		Rumus Entropy						
	0.75	0.25	0.811278124						
Hitungan Reminder									
	Jumlah Total	Jumlah Jenis Mawar	Jumlah Jenis Mawar	Proporsi Jenis Mawar	Proporsi jenis Mawar	Entropi	Pengali	Ent * Pengali	
Ukuran Batang									
Pendek	1	0	1	0	1	0	0.25	0	
Sedang	2	2	0	1	0	0	0.5	0	
Tinggi	1	1	0	1	0	0	0.25	0	
					Reminder			0	

Gambar 4 11 Perhitungan Keseluruhan Warna Kelopak Merah

Gambar diatas merupakan hasil perhitungan Information Gain pada atribut Ukuran_Batang yang dimana atribut tersebut digunakan ketika Warna_Kelopak = Merah sebagai berikut :

- Warna_Kelopak = Merah.
 - Ukuran_Batang Pendek = Mawar B.
 - Ukuran_Batang Sedang = Mawar A.
 - Ukuran_Batang Tinggi = Mawar B.
- Warna_Kelopak = Pink.

	A	B	C
1	Warna Kelopak	Ukuran Batang	Jenis Mawar
2	Pink	Sedang	Mawar A
3	Pink	Pendek	Mawar B
4	Pink	Tinggi	Mawar A
5	Pink	Tinggi	Mawar A
6			

Gambar 4 12 Dataset Warna Kelopak Pink

Berdasarkan gambar diatas dataset tersebut memiliki jumlah 4 data, diperoleh :

- Jenis Mawar A = 3.
- Jenis Mawar B = 1.

Sehingga entropy total dataset mawar adalah :

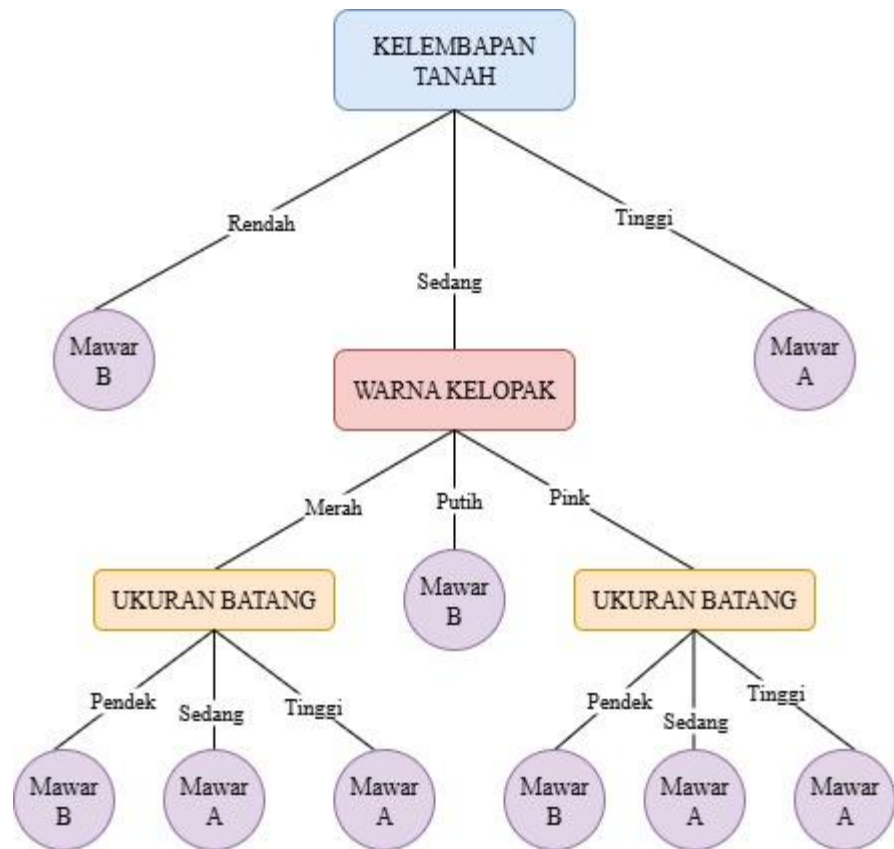
Entropi									
	Entropi	Reminder	IG						
IG Ukuran Batang	0.811278124	0	0.811278124						
Rincian Hitung Entropi									
Total Data	4								
Jenis Tulip Mawar A	3								
Jenis Tulip Mawar B	1								
	Proporsi Jenis A	Proporsi Jenis B	Rumus Entropy						
	0.75	0.25	0.811278124						
Hitungan Reminder									
Warna Kelopak									
	Jumlah Total	Jumlah Jenis Mawar	Jumlah Jenis Mawar	Proporsi Jenis Mawar	Proporsi Jenis Mawar	Entropi	Pengali	Ent * Pengali	
Ukuran Batang									
Pendek	1	0	1	0	1	0	0.25	0	
Sedang	1	1	0	1	0	0	0.25	0	
Tinggi	2	2	0	1	0	0	0.5	0	
					Reminder				0

Gambar 4 13 Perhitungan Keseluruhan Warna Kelopak Pink

Gambar diatas merupakan hasil perhitungan Information Gain pada atribut Ukuran_Batang yang dimana atribut tersebut digunakan ketika Warna_Kelopak = Pink sebagai berikut :

- Warna_Kelopak = Pink.
 - Ukuran_Batang Pendek = Mawar B.
 - Ukuran_Batang Sedang = Mawar A.
 - Ukuran_Batang Tinggi = Mawar A.

4) Membangun cabang Pohon Keputusan.



Gambar 4 14 Pohon Keputusan Mawar

Dari struktur pohon keputusan diatas dapat disimpulkan bahwa :

- Kelembapan_Tanah merupakan faktor yang berpengaruh dalam menentukan jenis mawar.
- Warna_Kelopak merupakan faktor tambahan khususnya pada atribut Kelembapan_Tanah = Sedang.
- Ukuran_Batang berperan pada kondisi tertentu, khususnya ketika pada atribut Warna_Kelopak masih belum bisa ditentukan jenisnya terutama pada Warna_Kelopak = Merah dan Warna_Kelopak = Pink.

b. Dataset Resiko Diabetes

- Menghitung Entropy

Rincian Hitung Entropi					
Total Data	26				
Jenis Rendah	4				
Jenis Sedang	12				
Jenis Tinggi	10				
		Proporsi Jeni	Proporsi Jeni	Proporsi Jeni	Rumus Entropy
		0.153846154	0.46153846	0.384615385	1.460484681

Gambar diatas merupakan langkah pertama dalam penerapan algoritma C4.5 untuk menghitung nilai entropy awal dari dataset mawar. Entropy ini digunakan untuk mengukur sebuah ketidakpastian atau keragaman pada data. Berdasarkan dataset tersebut memiliki sebanyak 26 data, diperoleh :

- Jumlah Rendah = 4.
- Jumlah Sedang = 12.
- Jumlah Tinggi = 10.

Sehingga entropy total dataset resiko diabetes 1.460.

- Menghitung Informasi Gain (IG) untuk tiap atribut.

Setelah entropy total diperoleh, langkah selanjutnya adalah menghitung nilai Informasi Gain (IG) dari masing-masing atribut. Nilai IG menunjukan seberapa besar atribut tersebut mampu mengurangi ketidakpastian entropy dan menjadi dasar untuk pemilihan node terbaik pada sebuah decision tree. Berikut hasil perhitungan Informtion Gain sebagai berikut :

Hitungan Reminder										
Usia										
	Jumlah Total	Jumlah Jenis	Jumlah Jenis	Jumlah Jenis	Proporsi Jenis	Proporsi jenis	Proporsi jenis	Entropi	Pengali	Ent * Pengali
Muda	7	4	2	1	0.571428571	0.285714286	0.142857	1.378783493	0.269231	0.371211
Dewasa	10	2	5	3	0.2	0.5	0.3	1.485475297	0.384615	0.571337
Tua	9	0	3	6	0	0.333333333	0.666667	0.918295834	0.346154	0.317872
					Reminder					1.260419
IMT										
Normal	9	4	4	1	0.444444444	0.444444444	0.111111	1.392147224	0.346154	0.481897
Obesitas	8	0	2	6	0	0.25	0.75	0.811278124	0.307692	0.249624
Overweight	9	0	6	3	0	0.666666667	0.333333	0.918295834	0.346154	0.317872
					Reminder					1.049393
Aktivitas Fisik										
Rendah	8	0	2	6	0	0.25	0.75	0.811278124	0.307692	0.249624
Sedang	9	1	5	3	0.111111111	0.555555556	0.333333	1.351644115	0.346154	0.467877
Tinggi	9	3	5	1	0.333333333	0.555555556	0.111111	1.351644115	0.346154	0.467877
					Reminder					1.185378

- Menentukan atribut dengan Gain terbesar.

Berikut hasil dari perhitungan IG tiap atribut yang akan menghasilkan Gain Terbesar. Dapat kita lihat dibawah ini untuk IG Usia = 0.200, IG IMT = 0.411, dan IG Aktivitas Fisik = 0.275. Maka dari itu kita dapat menyimpulkan bahwa IG terbesar yaitu IG IMT, sehingga IMT merupakan nood akar.

Entropi			
	Entropi	Reminder	IG
IG Usia	1.460484681	1.26041923	0.200065453
IG IMT	1.460484681	1.04939279	0.411091892
IG Aktivitas Fisik	1.460484681	1.18537766	0.275107025

- Dataset IMT = Normal.

Usia	Aktivitas Fisik	Risiko Diabetes
Dewasa	Rendah	Sedang
Tua	Rendah	Tinggi
Muda	Sedang	Rendah
Dewasa	Sedang	Sedang
Tua	Sedang	Sedang
Muda	Tinggi	Rendah
Dewasa	Tinggi	Rendah
Tua	Tinggi	Sedang
Dewasa	Tinggi	Rendah

Berdasarkan gambar diatas merupakan dataset lanjutan dari IMT Normal untuk mengklasifikasikan jenis resiko diabetes berdasarkan Usia dan Aktifitas Fisik , yaitu :

- Usia (Muda = 2, Dewasa = 4, Tua = 3).
- Aktifitas Fisik (Rendah = 2, Sedang = 3, Tinggi = 4).
- Dari label Resiko Diabetes (Rendah = 4, Sedang = 4, Tinggi = 1).
- Data tersebut berjumlah 9 baris data.

- Entropy Normal

Entropy yang dihasilkan adalah 0.139.

Rincian Hitung Entropi					
Total Data	9				
Jenis Rendah	4				
Jenis Sedang	4				
Jenis Tinggi	1				
	Proporsi J	Proporsi J	Proporsi J	Rumus Entropy	
	0.444444	0.444444	0.111111	1.392147	

- Information Gain tiap atribut.

Hitungan Reminder										
Usia										
	Jumlah To	Jumlah Je	Jumlah Je	umlah Jen	Proporsi J	Proporsi j	Proporsi j	Entropi	Pengali	Ent * Pengali
Muda	2	2	0	0	1	0	0	0	0.222222	0
Dewasa	4	2	2	0	0.5	0.5	0	1	0.444444	0.444444
Tua	3	0	2	1	0	0.666667	0.333333	0.918296	0.333333	0.306099
						Reminder				0.750543
Aktivitas Fisik										
Rendah	2	0	1	1	0	0.5	0.5	1	0.222222	0.222222
Sedang	3	1	2	0	0.333333	0.666667	0	0.918296	0.333333	0.306099
Tinggi	4	3	1	0	0.75	0.25	0	0.811278	0.444444	0.360568
						Reminder				0.888889

- Information Gain terbesar.

Entropi			
	Entropi	Reminder	IG
IG Usia	1.392147	0.750543	0.641604
IG Aktivitas Fisik	1.392147	0.888889	0.503258

Gambar diatas merupakan hasil penjumlahan dari aktivitas sebelumnya yakni dari penjumlahan Information Gain tiap atribut yang dimana terdiri dari :

- IG Usia : Entropy = 1.392, Reminder = 0.750, dan IG = 0.641.
- IG Aktivitas Fisik : Entropy = 1.392, Reminder = 0.888, dan IG = 0.503.

Dapat kita simpulkan bahwa information gain terbesar adalah IG Usia = 0.641.

- Dataset IMT = Obesitas.

Usia	Aktivitas Fisik	Risiko Diabetes
Muda	Rendah	Tinggi
Muda	Sedang	Sedang
Dewasa	Tinggi	Sedang
Dewasa	Sedang	Tinggi
Dewasa	Rendah	Tinggi

- Entropy Obesitas

Rincian Hitung Entropi				
Total Data	5			
Jenis Rendah	0			
Jenis Sedang	2			
Jenis Tinggi	3			
	Proporsi J	Proporsi J	Proporsi J	Rumus Entropy
	0	0.4	0.6	0.970951

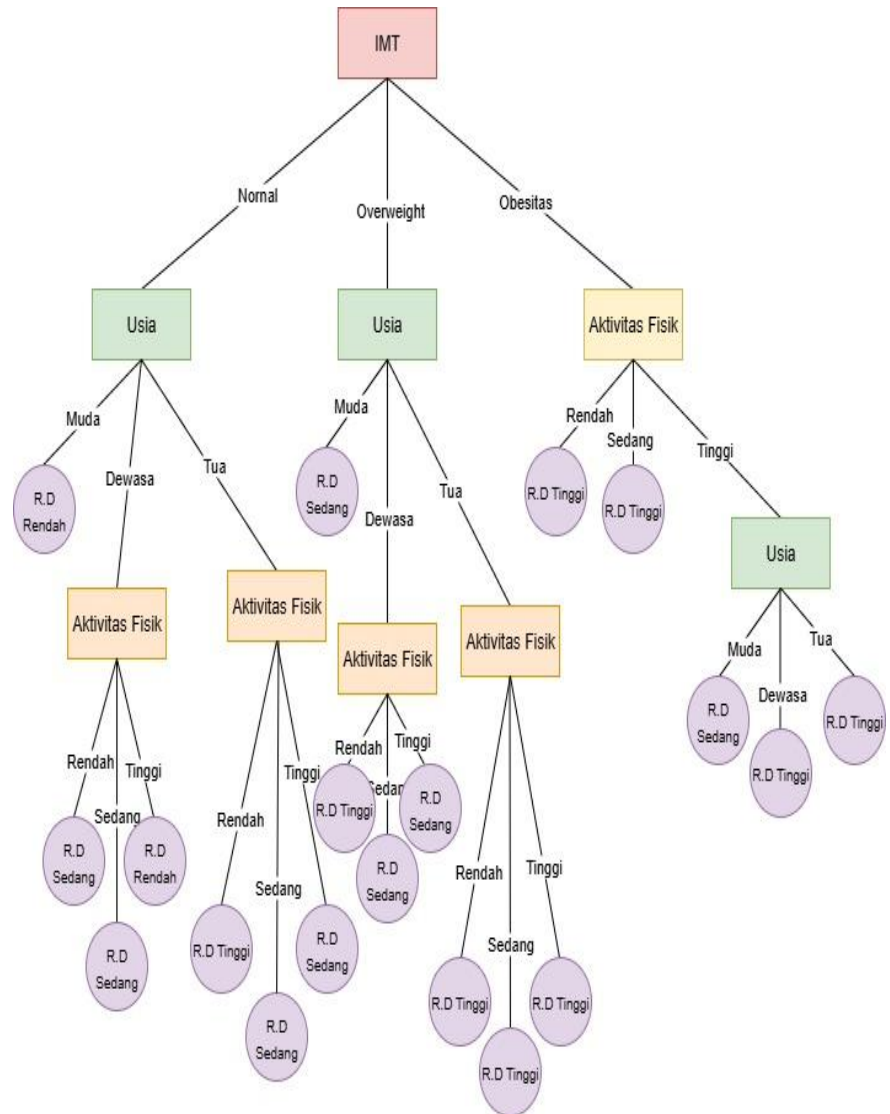
- Information Gain tiap atribut.

Hitungan Reminder											
Usia											
	Jumlah To	Jumlah Je	Jumlah Je	umlah Jer	Proporsi J	Proporsi j	Proporsi j	Entropi	Pengali	Ent * Pengali	
Muda	2	0	1	1	0	0.5	0.5	0	0.4	0	
Dewasa	3	0	1	2	0	0.333333	0.666667	0.918296	0.6	0.550978	
						Reminder				0.550978	
Aktivitas Fisik											
Rendah	2	0	0	2	0	0	1	0	0.4	0	
Sedang	2	0	1	1	0	0.5	0.5	1	0.4	0.4	
Tinggi	1	0	1	0	0	1	0	0	0.2	0	
						Reminder				0.4	

- Dataset IMT = Overweight.

Usia	Aktivitas Fisik	Risiko Diabetes
Muda	Tinggi	Sedang
Muda	Rendah	Sedang
Dewasa	Tinggi	Sedang
Dewasa	Sedang	Sedang
Dewasa	Rendah	Tinggi
Tua	Tinggi	Sedang
Tua	Sedang	Tinggi
Tua	Rendah	Tinggi
Muda	Sedang	Sedang

- Entropy Overweight.



4.3 Hasil Simulasi Perhitungan di ChatGPT Data Analytic

Dataset yang sama diberikan ke GPT Data Analytic dengan perintah (prompting) untuk menghitung entropy, information gain, dan buat decision tree C4.5 dataset berikut :

- a. Dataset Mawar

Langkah pertama dalam algoritma C4.5 adalah menghitung **Entropy** awal dataset:

1. Hitung jumlah kelas

Dari tabel:

- Mawar A = 11
- Mawar B = 14
- Total = 25

Data diatas merupakan hasil penjumlahan yang dilakukan oleh ChatGPT yang dimana jumlah total data yaitu 25. Dari tabel tersebut dapat dilihat total jumlah Mawar A = 11 dan Mawar B 14. Hasil ini memiliki perbedaan penjumlahan dengan hasil di Excel yang dimana penjumlahan di Excel Mawar A = 13 dan Mawar B 12.

1) Hitung Entropy

2. Rumus Entropy

$$Entropy(S) = - \sum p_i \log_2(p_i)$$

dengan p_i adalah proporsi tiap kelas.

3. Hitung Entropy awal dataset

$$\begin{aligned} p(A) &= \frac{11}{25} = 0.44, & p(B) &= \frac{14}{25} = 0.56 \\ Entropy(S) &= -(0.44 \cdot \log_2(0.44) + 0.56 \cdot \log_2(0.56)) \\ &= -(0.44 \cdot -1.185 + 0.56 \cdot -0.836) \\ &= 0.521 + 0.468 = 0.989 \end{aligned}$$

👉 Jadi **Entropy awal dataset = 0.989** (mendekati 1, artinya data cukup tidak pasti/bercampur).

Gambar diatas merupakan hasil penjumlahan data yang dihasilkan oleh ChatGPT yang dimana entropy awal dataset tersebut adalah 0.989. Untuk perhitungan data di Excel dengan ChatGPT itu untuk perhitungannya sedikit ada perbedaan. Pada Excel jumlah dataset untuk nilai Mawar A = 13 dan Mawar B = 12.

Rekap:

- Mawar A = 13 (baris ke-8, 9, 12, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25)
 - Mawar B = 12 (sisanya)
 - Total = 25 ✓
-

Jadi memang benar, hasil Anda lebih tepat:

- $p(A) = \frac{13}{25} = 0.52$
- $p(B) = \frac{12}{25} = 0.48$

Setelah melakukan perubahan jumlah data yang dihasilkan oleh ChatGPT dapat dilihat sebagai berikut :

Entropy awal (koreksi):

$$\begin{aligned}
 Entropy(S) &= -(0.52 \cdot \log_2(0.52) + 0.48 \cdot \log_2(0.48)) \\
 &= -(0.52 \cdot -0.943 + 0.48 \cdot -1.057) \\
 &= 0.490 + 0.507 = 0.997
 \end{aligned}$$

👉 Jadi Entropy awal dataset = 0.997

Gambar diatas merupakan hasil perubahan yang dilakukan oleh ChatGPT yang dimana menghasilkan entropy awal dataset adalah 0.997. Adapun perbedaan hasil dengan Excel yang dimana entropy awalnya adalah 0.998. Hal tersebut sangatlah wajar terjadi karena yang dihasilkan oleh ChatGPT menggunakan pembulatan 3 desimal.

2) Hitung Information Gain tiap atribut.

a) Warna Kelopak

- Merah (9): $p(A) = 5/9$, $p(B) = 4/9$
 $H = 0.9911$, $w = 9/25 = 0.36 \rightarrow 0.3568$
- Putih (8): $p(A) = 5/8$, $p(B) = 3/8$
 $H = 0.9544$, $w = 0.32 \rightarrow 0.3054$
- Pink (8): $p(A) = 3/8$, $p(B) = 5/8$
 $H = 0.9544$, $w = 0.32 \rightarrow 0.3054$

$$H(S \mid \text{Warna}) = 0.3568 + 0.3054 + 0.3054 = 0.9676$$

$$IG(\text{Warna}) = 0.9988 - 0.9676 = \mathbf{0.0312}$$

b) Ukuran Batang

- Pendek (6): (A=0, B=6) $\rightarrow$ pure B $\rightarrow H = 0$, *kontribusi* = 0
- Sedang (10): (A=6, B=4) $\rightarrow$
 $H = 0.9709$, $w = 0.4 \rightarrow 0.3884$
- Tinggi (9): (A=6, B=3) $\rightarrow$
 $H = 0.9183$, $w = 0.36 \rightarrow 0.3306$

$$H(S \mid \text{Ukuran}) = 0.7189$$

$$IG(\text{Ukuran}) = 0.9988 - 0.7189 = \mathbf{0.2799}$$

c) Kelembapan Tanah

- Rendah (7): (A=0, B=7) $\rightarrow$ pure B $\rightarrow H = 0$, *kontribusi* = 0
- Sedang (11): (A=6, B=5) $\rightarrow$
 $H = 0.9940$, $w = 0.44 \rightarrow 0.4373$
- Tinggi (7): (A=7, B=0) $\rightarrow$ pure A $\rightarrow H = 0$, *kontribusi* = 0

$$H(S \mid \text{Kelembapan}) = 0.4373$$

$$IG(\text{Kelembapan}) = 0.9988 - 0.4373 = \mathbf{0.5615}$$

3) Hitung Information terbesar.

- $IG(\text{Warna Kelopak}) = 0.0312$
- $IG(\text{Ukuran Batang}) = 0.2799$
- $IG(\text{Kelembapan Tanah}) = 0.5615$

👉 Jadi sama dengan logika Excel Anda: atribut terbaik untuk root adalah **Kelembapan Tanah**.

Gambar diatas merupakan hasil perhitungan dari data sebelumnya yang dimana $IG \text{Warna_Kelopak} = 0.0312$, $IG \text{Ukuran_Batang} = 0.2799$, dan $IG \text{Kelembapan_Tanah} = 0.5615$.

Dari hasil tersebut Information Gain paling besar adalah atribut $\text{Kelembapan_Tanah} = 0.5615$. Oleh karena itu, pada proses pembangunan pohon keputusan (Decision Tree) adalah Kelembapan_Tanah dipilih sebagai nood akar (root node). Hanya ada sedikit perbedaaan hasil penjumlahan IG terbesar pada Excel.

Pada gambar sebelumnya dalam perhitungan Information Gain masing-masing atribut kita dapat melihat sebagai berikut :

- Kelembapan_Tanah rendah, maka keputusan langsung menunjukan bahwa jenis mawar adalah Mawar B.
- Kelembapan_Tanah tinggi, maka keputusan langsung menunjukan bahwa jenis mawar adalah Mawar A.
- Jika Kelembapan_Tanah berada pada kondisi sedang, maka klasifikasi tidak dapat ditentukan sehingga diperlukannya atribut lain sebagai akar selanjutnya. Dikarenakan data masih berbeda dan belum bisa menentukan jenis mawar.

Maka dari itu kita menghitung ulang dari awal dengan data dari $\text{Kelembapan_Tanah} = \text{Sedang}$ sebagai berikut :

- Hitung Entropy $\text{Kelembapan_Tanah} = \text{Sedang}$.

$$H(S_{\text{sedang}}) = - \frac{6}{11} \log_2 \frac{6}{11} - \frac{5}{11} \log_2 \frac{5}{11} = 0.9940302115.$$
- Hitung IG tiap atribut.
 - Warna_Kelopak (Subset = 11 baris).
 Distribusi dalam subset Sedang:
 - Merah : total 4 $\rightarrow A = 3, B = 1$
 - Putih : total 3 $\rightarrow A = 0, B = 3$

- Pink : total 4 $\rightarrow A = 3, B = 1$

Per-node entropi:

- $H(\text{Merah}) = H(3,1) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} = 0.8112781245.$
- $H(\text{Putih}) = H(0,3) = 0.0$ (murni).
- $H(\text{Pink}) = H(3,1) = 0.8112781245.$

Bobot & kontribusi tiap nilai :

- Merah : Bobot = 4/11. Kontribusi = (4/11) x 0.8112781245 = 0.295037.
- Putih : Bobot = 3/11. Kontribusi = (3/11) x 0 = 0.0.
- Pink : Bobot = 4/11. Kontribusi = (4/11) x 0.8112781245 = 0.295037.

Jumlah entropi ter-kondisi :

- $H(\text{Ssedang} \mid \text{Warna_Kelopak}) = 0.295037 + 0 + 0.295037 = 0.590074.$

Information Gain untuk Warna_Kelopak :

- $IG(\text{Warna_Kelopak}) = H(\text{Ssedang}) - H(\text{Ssedang} \mid \text{Warna_Kelopak}) = 0.9940302 - 0.590074 = 0.4040.$

- Ukuran_Batang (Subset = 11 baris).

Distribusi dalam subset Sedang:

- Pendek : total 3 $\rightarrow A = 0, B = 3.$
- Sedang : total 4 $\rightarrow A = 3, B = 1.$
- Tinggi : total 4 $\rightarrow A = 3, B = 1,$

Per-node entropi:

- $H(\text{Pendek}) = H(0,3) = 0.0.$
- $H(\text{Sedang}) = H(3,1) = 0.8112781245.$
- $H(\text{Tinggi}) = H(3,1) = 0.8112781245.$

Bobot & kontribusi tiap nilai :

- Pendek : Bobot = 3/11. Kontribusi = (3/11) x 0 = 0.0.
- Sedang : Bobot = 4/11. Kontribusi = (4/11) x 0.8112781245 = 0.295037.

- Tinggi : Bobot = $4/11$. Kontibusi = $(4/11) \times 0.8112781245 = 0.295037$.

Jumlah entropi ter-kondisi :

- $H(\text{Ssedang} \mid \text{Ukuran_Batang}) = 0 + 0.295037 + 0.295037 = 0.590074$.

Information Gain untuk Warna_Kelopak :

- $IG(\text{Ukuran_Batang}) = 0.9940302 - 0.590074 = 0.4040$.

c) Hitung IG terbesar.

- $IG(\text{Warna_Kelopak}) = H(\text{Ssedang}) - H(\text{Ssedang} \mid \text{Warna}) = 0.9940302 - 0.590074 = 0.4040$.
- $IG(\text{Ukuran_Batang}) = 0.9940302 - 0.590074 = 0.4040$.

Pada subset Kelembapan_Tanah = Sedang, $IG(\text{Warna_Kelopak}) = IG(\text{Ukuran_Batang}) = 0.4040$ (nilai sama/tie). Yang artinya Warna_Kelopak ataupun Ukuran_Batang sama-sama baik sebagai atribut pembagi berikutnya.

Untuk membuat cabang dari subset lanjutan disini mengambil dari Kelembapan_Tanah = Sedang dengan node lanjutan = Warna_Kelopak. Kita sudah tahu di subset ini :

- Total data = 11 (A = 6, B = 5).
- Entropy parent = 0.994.
- Atribut terbaik setelah tie/sama kita pilih Warna_Kelopak.

1) Cabang Warna_Kelopak (pada subset Kelembapan_Tanah = Sedang).

- Warna_Kelopak = Merah (4 data, A = 3, B = 1).
 - Entropy = 0.811
 - Masih campur (tidak murni).
 - Perlu dicek **atribut sisa = Ukuran Batang** untuk split lebih lanjut.

Distribusi Ukuran Batang pada Warna Merah & Kelembapan Sedang:

- Sedang = 2 $\rightarrow$ (A=1, B=1) $\rightarrow$ Entropy = 1.0 (maks ketidakpastian).

- Tinggi = 2 $\rightarrow$ (A=2, B=0) $\rightarrow$ Entropy = 0 (murni A)

Jika di-split:

- Untuk Ukuran_Batang = Tinggi $\rightarrow$ semua A.
 - Untuk Ukuran_Batang = Sedang $\rightarrow$ campur (1 A, 1 B) $\rightarrow$ biasanya di-stop (karena datanya sangat kecil) atau diputuskan mayoritas.
- Warna_Kelopak = Putih (3 data, A = 0, B = 3).
 - Entropy = 0.0 (murni).
 - Jadi cabang ini langsung Mawar B.
 - Warna_Kelopak = Pink (4 data, A = 0, B = 1).
 - Sama seperti Warna_Kelopak Merah.

Distribusi Ukuran_Batang :

- Sedang = 2 (A = 1, B = 1), entropy 1.0.
- Tinggi = 2 (A = 2, B = 0), entropy = 0.

Hasil split :

- Ukuran_Batang = Tinggi, semua A.
- Ukuran_Batang = Sedang, campur (1 A, 1 B) : biasanya diputuskan mayoritas atau dianggap node terakhir.

2) Ringkasan Tree sampai tahap ini.

- Root = Kelembapan_Tanah
 - Rendah : Mawar B (murni).
 - Tinggi : Mawar A (murni).
 - Sedang : lanjut split Warna_Kelopak.
 - Merah.
 - Ukuran_Batang = Tinggi, Mawar A.
 - Ukuran_Batang = Sedang, (1 A, 1 B), mayoritas? (bisa pilih A atau B. tergantung aturan stop).
 - Putih : Mawar B (murni)

- Pink.
 - Ukuran_Batang = Tinggi, Mawar A.
 - Ukuran_Batang = Sedang, (1 A, 1 B), mayoritas (A atau B)

Pertanyaan :

Kita berada di subset Kelembapan_Tanah = Sedang (11 data, A = 6, B = 5), lalu kita pilih Warna_Kelopak sebagai node.

- 1) Warna_Kelopak = Merah (4 data, A = 3, B = 1).

Distribusi menurut Ukuran_Batang :

- Pendek = 1, (A = 0, B = 1), entropy = 0.
- Sedang = 2, (A = 2, B = 0), entropy 0.
- Tinggi = 1, (A = 1, B = 0), entropy 0.

Semua cabang sudah murni, jadi :

- Pendek, Mawar B.
- Sedang, Mawar A.
- Tinggi, Mawar A.

- 2) Warna_Kelopak = Putih (3 data, A = 0, B = 3).

- Murni, semua Mawar B.

- 3) Warna_Kelopak = Pink (4 data, A = 3, B = 1).

Distribusi menurut Ukuran_Batang :

- Pendek = 1, (A = 0, B = 1), entropy = 0.
- Sedang = 1, (A = 1, B = 0), entropy = 0.
- Tinggi = 2, (A = 2, B = 0), entropy 0.

Semua cabang sudah murni, jadi :

- Pendek, Mawar B.
- Sedang, Mawar A.
- Tinggi, Mawar A.

- 4) Ringkasan Pohon Keputusan (sampai selesai).

- Root : Kelembapan_Tanah.
 - Rendah, Mawar B.
 - Tinggi, Mawar A.

- Sedang, lanjut split dengan Warna_Kelopak.
 - Merah.
 - Ukuran_Batang Pendek, Mawar B.
 - Ukuran_Batang Sedang, Mawar A.
 - Ukuran_Batang Tinggi, Mawar A.
 - Putih, Mawar B.
 - Pink.
 - Ukuran_Batang Pendek, Mawar B.
 - Ukuran_Batang Sedang, Mawar A.
 - Ukuran_Batang Tinggi, Mawar A.

4) Membangun cabang Pohon Keputusan



b. Dataset Resiko Diabetes

Dataset ini berisi 26 baris data dengan atribut predator :

- Usia (Muda, Dewasa, Tua).
- IMT (Normal, Overweight, Obesitas).
- Aktivitas Fisik (Tinggi, Sedang, Rendah).

Dan kelas targetnya : Resiko Diabetes (Rendah, Sedang, Tinggi).

Mulai dengan langkah awal dalam Algoritma Decision Tree C4.5, yaitu menghitung entropy awal dari dataset tersebut sebagai berikut :

- Resiko Rendah = 6 data.
- Resiko Sedang = 11 data.
- Resiko Tinggi = 9 data.
- Total = 26 data.

Sehingga proporsi yang dihasilkan adalah :

- $P_{Rendah} = 6/26 = 0.231$.
- $P_{Sedang} = 11/26 = 0.423$.
- $P_{Tinggi} = 9/26 = 0.346$.

Hitung entropy awal:

$$\text{Entropy}(S) = - (0.231 \cdot \log_2(0.231) + 0.423 \cdot \log_2(0.423) + 0.346 \cdot \log_2(0.346)) = 1.53.$$

Jadi Entropy awal dataset yaitu = 1.53.

Menghitung Gain untuk tiap atribut (Usia, IMT, Aktivitas Fisik) agar bisa menentukan root node pohon keputusan.

Ada sedikit perbedaan pencatatan jumlah kelas target. Maka dari itu kita menghitung ulang berdasarkan jumlah yang ditentukan :

- Total data = 26.
- Resiko Rendah = 4.
- Resiko Sedang = 12.
- Resiko Tinggi = 10.

Hitung proporsi :

- $P_{Rendah} = 4/26 = 0.154$.

- $PSedang = 12/26 = 0.462$.
- $PTinggi = 10/26 = 0.385$.

Hitung Entropy awal dataset.

$$\begin{aligned}
 Entropy(S) &= -(p_{Rendah} \log_2 p_{Rendah} + p_{Sedang} \log_2 p_{Sedang} + p_{Tinggi} \log_2 p_{Tinggi}) \\
 &= -(0.154 \cdot \log_2(0.154) + 0.462 \cdot \log_2(0.462) + 0.385 \cdot \log_2(0.385)) \\
 &\approx 1.38
 \end{aligned}$$

Gambar diatas merupakan hasil dari perhitungan entropy yang saya masukan sehingga entropy awal dataset = 1.38. Tetapi ada perbedaan dari hitung proporsi yang dihasilkan oleh ChatGPT dan Excel. Maka dari itu ChatGPT menghitung ulang lagi entropy awal dataset dengan proporsi (0.1538, 0.4615, 0.3846) sehingga menghasilkan sebagai berikut :

1) Entropy awal dataset.

$$H(S) = -(\sum p \log_2 p) \approx 1.4605$$

2) Information Gain tiap atribut (berdasarkan 26 baris data).

a) Usia.

- Muda : (R = 2, S = 4, T = 1) (n = 7).
 - Dewasa : (R = 2, S = 5, T = 3) (n = 10).
 - Tua : (R = 0, S = 3, T = 6) (n = 9).
- Reminder = 1.2604, Gain = 0.2001.

b) IMT.

- Normal : (R = 4, S = 4, T = 1) (n = 9).
 - Overweight : (R = 0, S = 6, T = 3) (n = 9).
 - Obesitas : (R = 0, S = 2, T = 6) (n = 8).
- Reminder = 1.0494, Gain = 0.4111.

c) Aktivitas Fisik.

- Tinggi : (R = 3, S = 5, T = 1) (n = 9).
- Sedang : (R = 1, S = 5, T = 3) (n = 9).
- Rendah : (R = 0, S = 2, T = 6) (n = 8).

Reminder = 1.1854, Gain = 0.2751.

3) Information Gain terbesar.

Kesimpulan (root) atribut dengan Gain terbesar adalah IMT = 0.4111, menjadikan root node.

✚ Root Node : IMT.

✚ Cabang 1 : IMT = Normal.

Distribusi kelas = (Rendah = 4, Sedang = 4, Tinggi = 1).

Entropy = 1.3921.

Hasil split terbaik = usia (Gain = 0.6416).

- Muda : (R = 2), daun : Rendah.
- Dewasa : (R = 2, S = 2), entropy = 1 perlu cek Aktivitas Fisik.
- Tua : (S = 2, T = 1), mayoritas Sedang, bisa dicek lagi dengan Aktivitas Fisik.

✚ Cabang 2 : IMT = Overweight.

Distribusi kelas = (Sedang = 6, Tinggi = 3).

Entropy = 0.9183.

Split Usia dan Aktivitas Fisik memberi Gain sama (0.3061).

Kita bisa pilih salah satu, misalnya Usia.

- Muda : (S = 3), daun : Sedang.
- Dewasa : (S = 2, T = 1), mayoritas Sedang.
- Tua : (T = 2, S = 1), mayoritas Tinggi.

✚ Cabang 3 : IMT = Obesitas.

Distribusi kelas = (Tinggi = 6, Sedang = 2).

Entropy = 0.8113.

Split Usia atau Aktivitas Fisik memberi Gain = 0.2169.

Pilih Aktivitas Fisik :

- Rendah : (T = 3), daun : Tinggi.
- Sedang : (T = 2, S = 1), mayoritas Tinggi.
- Tinggi : (S = 1, T = 1), masih tercampur, entropy = 1, bisa cek Usia lagi.

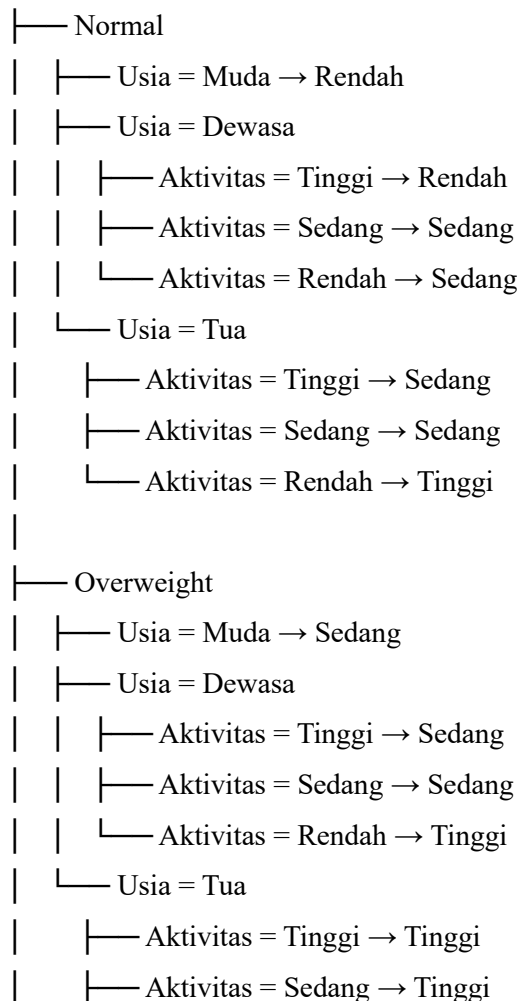
✚ Ringkasan sementara :

Root = IMT.

- Normal – split Usia.
 - Muda – Rendah.
 - Dewasa – (perlu split lagi Aktivitas Fisik).
 - Tua – mayoritas Sedang.
- Overweight – split Usia.
 - Muda – Sedang.
 - Dewasa – mayoritas Sedang.
- Obesitas – split Aktivitas Fisik.
 - Rendah – Tinggi.
 - Sedang – mayoritas Tinggi.
 - Tinggi – (Sedang/Tinggi), bisa split lagi

➤ Pohon Keputusan (Decision Tree) – Dataset Resiko Diabetes.

IMT?



```

|      └─ Aktivitas = Rendah → Tinggi
|
|
└─ Obesitas
    └─ Aktivitas = Rendah → Tinggi
    └─ Aktivitas = Sedang → Tinggi
    └─ Aktivitas = Tinggi
        └─ Usia = Muda → Sedang
        └─ Usia = Dewasa → Tinggi
        └─ Usia = Tua → Tinggi

```

➤ Atruran IF-THEN (Rules).

1. IF IMT = Normal

- IF IMT = Normal AND Usia = Muda → Risiko = Rendah
- IF IMT = Normal AND Usia = Dewasa AND Aktivitas = Tinggi → Risiko = Rendah
- IF IMT = Normal AND Usia = Dewasa AND Aktivitas = Sedang → Risiko = Sedang
- IF IMT = Normal AND Usia = Dewasa AND Aktivitas = Rendah → Risiko = Sedang
- IF IMT = Normal AND Usia = Tua AND Aktivitas = Tinggi → Risiko = Sedang
- IF IMT = Normal AND Usia = Tua AND Aktivitas = Sedang → Risiko = Sedang
- IF IMT = Normal AND Usia = Tua AND Aktivitas = Rendah → Risiko = Tinggi

2. IMT = Overweight.

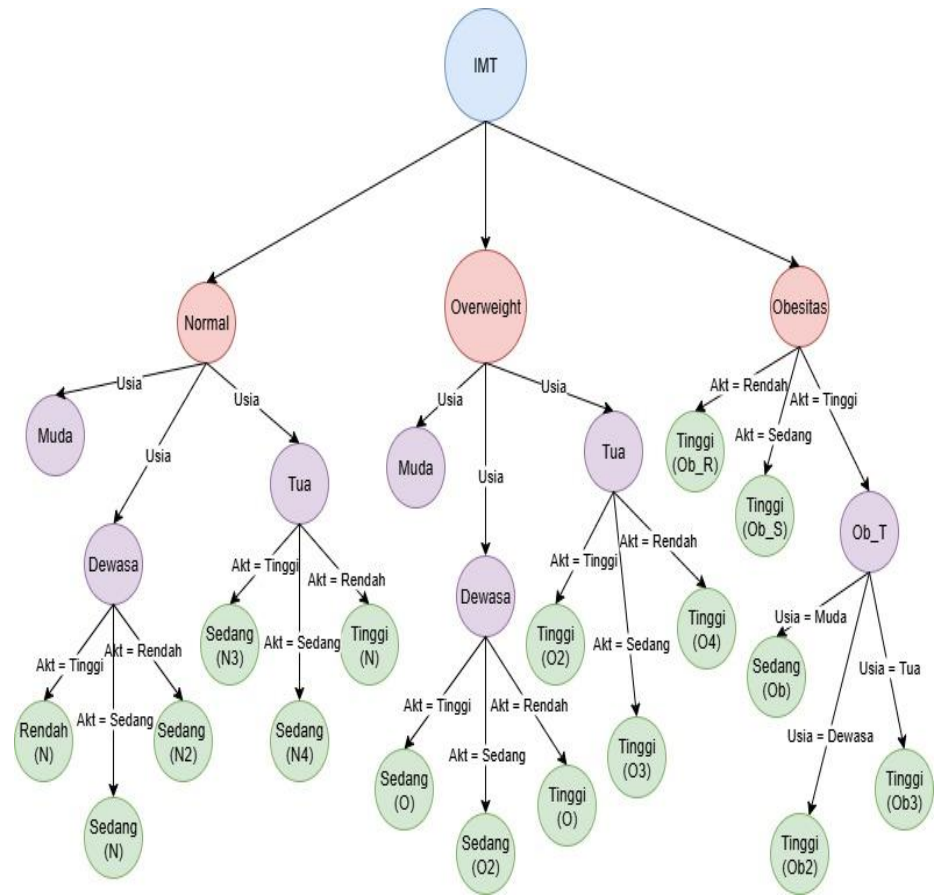
- IF IMT = Overweight AND Usia = Muda → Risiko = Sedang.
- IF IMT = Overweight AND Usia = Dewasa AND Aktivitas = Tinggi → Risiko = Sedang.
- IF IMT = Overweight AND Usia = Dewasa AND Aktivitas = Sedang → Risiko = Sedang.

- IF IMT = Overweight AND Usia = Dewasa AND Aktivitas = Rendah → Risiko = Tinggi.
- IF IMT = Overweight AND Usia = Tua AND Aktivitas = Tinggi → Risiko = Tinggi.
- IF IMT = Overweight AND Usia = Tua AND Aktivitas = Sedang → Risiko = Tinggi.
- IF IMT = Overweight AND Usia = Tua AND Aktivitas = Rendah → Risiko = Tinggi.

3. IMT = Obesitas.

- IF IMT = Obesitas AND Aktivitas = Rendah → Risiko = Tinggi
- IF IMT = Obesitas AND Aktivitas = Sedang → Risiko = Tinggi
- IF IMT = Obesitas AND Aktivitas = Tinggi AND Usia = Muda → Risiko = Sedang
- IF IMT = Obesitas AND Aktivitas = Tinggi AND Usia = Dewasa → Risiko = Tinggi
- IF IMT = Obesitas AND Aktivitas = Tinggi AND Usia = Tua → Risiko = Tinggi

➤ Decision Tree – Resiko Diabetes



4.4 Perbandingan Hasil Perhitungan

Bagian ini adalah hasil perbandingan yang telah dilakukan dalam proses yang pertama menggunakan Microsoft excel dan yang kedua menggunakan ChatGPT :

- a. Perbandingan Hasil Perhitungan Dataset Mawar.

Table 4.1 Perbandingan Hasil Perhitungan Dataset Mawar.

Tahap Perhitungan	Microsoft Excel	ChatGPT AI
Entropy (target)	0.998	0.997
Information Gain (Warna Kelopak)	0.0	0.0
Information Gain (Ukuran Batang)	0.279	0.279

Information Gain (Kelembapan Tanah)	0.981	0.561
Root ()	Kelembapan_Tanah	Kelembapan_Tanah

Data diatas menghasilkan Information Gain (Kelembapan Tanah) yang berbeda dikarenakan perhitungan excel menghitung berdasarkan data asli yaitu 12 A dan 13 B sedangkan ChatGPT menghitung setelah membaca ulang dataset melalui prompt yaitu 11 A dan 14 B. Rumus Information Gain yang sangat sensitive terhadap proporsi setiap kelas dalam subset, jika proporsi berubah sedikit saja maka entropy total berubah, entropy tiap subset berubah, serta IG ikut berubah meskipun hanya selisih 1 data.

b. Perbandingan Hasil Perhitungan Dataset Resiko Diabetes.

Table 4.2 Perbandingan Hasil Perhitungan Dataset Resiko Diabetes.

Tahap Perhitungan	Microsoft Excel	ChatGPT AI
Entropy (Target)	1.460	1.460
Information Gain (Usia)	0.200	0.200
Information Gain (IMT)	0.411	0.411
Information Gain (Aktivitas Fisik)	0.275	0.275
Root ()	IMT	IMT

4.5 Analisis Pengalaman Mahasiswa

Table 4.3 Ringkasan Hasil Kuesioner Pengalaman Mahasiswa

Aspek Penilaian	Pertanyaan	Pilih Jawaban	Hasil
Pemahaman Algoritma C4.5.	Sebelumnya mengikuti penelitian ini, apakah anda sudah pernah mempelajari Algoritma Decision Tree C4.5?	Belum pernah.	Mayoritas mahasiswa pernah, tetapi masih kurang pemahaman dalam Algoritma Decision Tree C4.5.
		Pernah, tapi masih kurang paham.	
		Sudah cukup paham.	

	Setelah mendapatkan penjelasan dan contoh, bagaimana tingkat pemahaman anda?	Masih bingung.	Cukup paham.
		Cukup paham.	
		Sangat paham.	
Pengalaman menggunakan Microsoft Excel.	Menurut anda, penggunaan Excel dalam simulasi Algoritma Decision Tree C4.5 :	Membantu memahami konsep.	Membantu dan transparan perhitungan terlihat jelas, tetapi rumit dan rawan salah input data.
		Transparan (perhitungan terlihat jelas)	
		Rumit dan memakan waktu.	
		Rentan salah input.	
Pengalaman menggunakan ChatGPT.	Menurut anda, penggunaan ChatGPT dalam simulasi Algoritma Decision Tree C4.5 :	Praktis dan cepat.	Praktis cepat, dan mudah untuk dipahami, tetapi masih perlu dilakukannya pengecekan ulang akurasi.
		Memberikan hasil pohon keputusan secara instan.	
		Penjelasan dengan bahasa yang sederhana dan mudah untuk dipahami.	
		Akurasinya masih perlu dicek secara manual.	
Perbandingan Excel dan ChatGPT	Menurut anda, manakah yang lebih efektif untuk dipelajari mahasiswa?	Excel lebih baik untuk pemahaman dasar.	Kombinasi keduanya lebih optimal, karena saling melengkapi.
		ChatGPT lebih baik untuk efisiensi dan praktis.	
		Kombinasi keduanya lebih optimal.	

BAB 5

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Berdasarkan hasil penelitian mengenai Analisis Perbandingan Hasil Dari Algoritma Decision Tree C4.5 Menggunakan Microsoft Excel Dan ChatGPT, maka dapat ditarik beberapa Kesimpulan sebagai berikut :

1. Mekanisme hasil kerja algoritma decision tree C4.5 dapat dipahami dengan baik melalui simulasi pada Microsoft Excel. Excel membantu mahasiswa dapat melihat proses perhitungan Entropy, Information Gain, dan Gain Ratio secara manual, sehingga memudahkan pemahaman konsep algoritma menjadi lebih jelas.
2. Penggunaan ChatGPT AI (Data Analytic) memungkinkan mahasiswa memperoleh hasil perhitungan Algoritma Decision Tree C4.5 secara cepat. ChatGPT tidak hanya memberikan hasil akhir, tetapi juga dapat menjelaskan secara detail tahap demi tahap untuk perhitungan dengan Bahasa yang lebih mudah dipahami dan interaktif.
3. Hasil perbandingan antara Algoritma Decision Tree C4.5 menggunakan microsoft excel dan ChatGPT menunjukkan bahwa adanya ketidaksamaan. Perbedaan ini tampak pada struktur pohon yang dihasilkan, jumlah node, serta akurasi klasifikasi. Selain itu, meskipun jumlah baris data yang digunakan sama pada kedua metode, hasil klasifikasi label pada beberapa baris data berbeda. Microsoft excel memberikan hasil yang lebih konsisten sesuai perhitungan manual langkah demi langkah (entropy dan gain), sedangkan ChatGPT menghasilkan pohon Keputusan lebih cepat namun dengan variasi pada penentu label akhir. Hal ini menunjukkan bahwa meskipun algoritma yang digunakan sama, proses perhitungan dan interpretasi sistem menghasilkan keluaran yang tidak sepenuhnya identik.
4. Perbedaan nilai Information Gain (IG) Kelembapan Tanah terjadi bukan karena algoritma C4.5 berbeda, tetapi karena input data yang dibaca oleh ChatGPT tidak 100% identik dengan dataset asli di Excel. Adapun perbedaan jumlah Mawar A dan B (Excel : 12/13 sedangkan ChatGPT : 11/14). Sehingga entropy awal berubah, IG tiap atribut berubah, termasuk Kelembapan Tanah juga ikut berubah.

5.2 SARAN

1. Untuk mahasiswa sebaiknya bisa memanfaatkan kedua metode : Excel untuk memahami dasar perhitungan Algoritma Decision Tree C4.5, dan untuk ChatGPT AI memudahkan simulasi serta latihan soal dengan menggunakan data yang lebih besar.
2. Untuk dosen, disarankan untuk mengkombinasikan pembelajaran tradisional (manual/excel) dengan pemanfaatan ChatGPT sebagai media pembelajaran interaktif, sehingga mahasiswa bisa mudah memahami teori sekaligus praktik dengan lebih efektif dan efisiensi waktu.
3. Untuk penelitian selanjutnya, dapat dilakukan pengujian dengan dataset yang lebih besar dan beragam, serta membandingkan algoritma decision tree c4.5 dengan algoritma klasifikasi lainnya. Misalnya menggunakan Random Forest, Naïve Bayes, atau ID3 menggunakan ChatGPT AI ini.

DAFTAR PUSTAKA

- [KCSIB18] Kurniawan, C., Sari, I., & Brata, D. (2018). Analysis of student demographic information using data mining classification with decision tree..
- [LZWT20] Liu, Z., Wen, T., Sun, W., & Zhang, Q. (2020). Semi-supervised self-training feature weighted clustering decision tree and random forest. *Ieee Access*, 8, 128337-128348
- [MM23] Misnawati, M. (2023). Chatgpt: keuntungan, risiko, dan penggunaan bijak dalam era kecerdasan buatan. *Prosiding Seminar Nasional Pendidikan Bahasa Sastra Seni Dan Budaya*, 2(1), 54-67.
- [AMA23] Abas, M., Arumugam, S., Yunus, M., & Rafiq, K. (2023). Chatgpt and personalized learning: opportunities and challenges in higher education. *International Journal of Academic Research in Business and Social Sciences*, 13(12).
- [CY23] Chen, Y. (2023). Influencing factor analysis for automobile customers based on one-way anova and decision tree. *Advances in Economics Management and Political Sciences*, 21(1), 356-364.
- [SRXS2] Shan, R., Xiao, X., Che, J., Du, J., & Li, Y. (2022). Data mining optimization software and its application in financial audit data analysis. *Mobile Information Systems*, 2022, 1-7.
- [SAK20] Suherman, A., Kurnaedi, D., Lusa, S., & Darmawan, R. (2020). Junior class preparedness classification faces a national exam using c.45 algorithm with a particle swarm optimization approach. *Bit-Tech*, 2(3), 101-109.
- [FSM24] Fatima, A., Shafique, M., Alam, K., Ahmad, T., & Mustafa, M. (2024). Chatgpt in medicine: a cross-disciplinary systematic review of chatgpt's (artificial intelligence) role in research, clinical practice, education, and patient interaction. *Medicine*, 103(32), e39250.
- [FTS23] Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A., & Hatta, H. (2023). Analysing the role of chatgpt in improving student productivity in higher education. *Journal on Education*, 5(4), 14886-14891.

- [FZ23] Fu, P. and Zhang, P. (2023). The issues and optimization strategies in the modernization of vocational education in the era of chat gpt. *World Journal of Educational Research*, 10(5), p251
- [KAZ24] Kharis, S., Arisanty, M., & Zili, A. (2024). Pengalaman dan perspektif pendidik terhadap penggunaan chatgpt dalam pengajaran. *Jurnal Pendidikan*, 33(1), 515-524.
- [NTK24] Nguyen-Trung, K. (2024). Chatgpt in thematic analysis: can ai become a research assistant in qualitative research?..