

---

**Analisis Minat Masyarakat Menggunakan Media Sosial Menggunakan  
Algoritma C4.5 dan Metode Naïve Bayes**

**Nia Putri Panjaitan<sup>1</sup>, Syaiful Zuhri Harahap<sup>2</sup>, Rahma Muti Ah<sup>3</sup>**

Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu<sup>1,2,3</sup>

Email : [niaputri1511@gmail.com](mailto:niaputri1511@gmail.com)<sup>1</sup>, [syaifulzuhriharahap@gmail.com](mailto:syaifulzuhriharahap@gmail.com)<sup>2</sup>,  
[rmuthea5@gmail.com](mailto:rmuthea5@gmail.com)<sup>3</sup>

Corresponding Author : [niaputri1511@gmail.com](mailto:niaputri1511@gmail.com)

**Abstract**

*The analysis of public interest using social media in data mining aims to understand user preferences and interests in various topics or products. By analyzing data from social media platforms, such as posts, comments, and interactions, researchers can identify significant interest patterns and trends, which can be used for more effective marketing strategies or product development that suits the public's desires. Common methods used in this analysis are the C4.5 and Naive Bayes algorithms. The C4.5 algorithm builds a decision tree that makes it easy to visualize and interpret the main factors that influence public interest. Meanwhile, Naive Bayes, with its probabilistic approach, classifies data based on existing features, providing fast and accurate predictions. Both methods are applied to process data from social media and produce in-depth insights into user preferences. The results of the analysis show that the prediction and classification of public interest have good accuracy, with the comparison result values showing very satisfactory performance. Both are able to identify and classify interests accurately, utilizing the advantages of each method to provide a better understanding of what is interesting to the public on social media.*

**Keywords:** Classification; C4.5 Algorithm; Naive Bayes Method.

**I. Pendahuluan**

Di era digital modern ini, Indonesia mengalami pergeseran besar dalam pola konsumsi informasi masyarakatnya. Meskipun sebagian besar orang masih menggunakan koran dan majalah sebagai sumber berita, semakin banyak individu yang beralih ke platform digital untuk memperoleh informasi terkini. Kemajuan teknologi, khususnya akses mudah ke internet dan perkembangan media sosial, telah menjadi katalisator utama perubahan

ini. Masyarakat Indonesia kini lebih cenderung mendapatkan berita melalui situs web berita, aplikasi berita, dan platform media sosial. Di Indonesia, media sosial yang paling banyak diminati oleh masyarakat umumnya adalah Instagram, Facebook, dan YouTube. Instagram menonjol karena popularitasnya sebagai platform berbagi foto dan video, terutama di kalangan generasi muda. Facebook tetap menjadi salah satu platform yang dominan dengan berbagai fitur seperti grup,

halaman bisnis, dan berbagi konten. Sementara itu, YouTube terus menjadi destinasi utama untuk konten video, baik hiburan, edukasi, maupun informasi. Keberagaman konten dan interaksi sosial yang ditawarkan oleh platform-platform ini membuat mereka tetap menjadi pilihan utama bagi pengguna media sosial di Indonesia. Facebook dan Twitter, sebagai platform berita sosial yang lebih tradisional, masih tetap relevan di Indonesia. Banyak media berita, organisasi pemerintah, dan individu menggunakan Facebook untuk membagikan artikel, video, dan informasi terkini. Twitter, dengan format cuitan yang singkat, menjadi tempat bagi berbagai diskusi, opini, dan pembaruan langsung dari berbagai sumber. Namun, perlu diingat bahwa sumber informasi di media sosial tidak selalu terverifikasi dengan baik, dan adanya penyebaran informasi palsu atau hoaks menjadi tantangan yang perlu diatasi. penulis ingin melakukan sebuah Penelitian tentang minat Masyarakat Indonesia pada media sosial. penelitian ini bertujuan untuk mendalami minat masyarakat Indonesia terhadap media sosial dengan pendekatan data mining menggunakan dua metode klasifikasi, yaitu Algoritma C4.5 dan Metode Naive Bayes. Pemilihan kedua metode ini didasarkan pada keunggulan mereka dalam mengolah dan menganalisis data besar untuk mengidentifikasi pola atau tren yang mungkin sulit dipahami secara manual. Algoritma C4.5 dikenal efektif dalam pembuatan pohon keputusan yang dapat mengelompokkan data berdasarkan fitur-fitur yang signifikan, sedangkan Metode Naive Bayes mengandalkan teorema Bayes untuk

memperkirakan probabilitas kelas tertentu berdasarkan fitur-fitur yang diberikan.

## II. Landasan Teori

### Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data.

### Naïve Bayes

Menurut Oslan dan Delen dalam (Kurniawan, Afif F, 2013) Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris Thomas Bayes yaitu dengan memprediksi masa depan berdasarkan data dari masa lalu. Naïve Bayes untuk setiap kelas keputusan menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vector informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan.

## III. Metode Penelitian

Algoritma C4.5 dan Naive Bayes adalah dua metode populer dalam data mining yang digunakan untuk analisis minat masyarakat pada media sosial. C4.5, yang merupakan perbaikan dari ID3, adalah algoritma

pohon keputusan yang bekerja dengan membagi data ke dalam subset yang lebih kecil berdasarkan nilai atribut yang paling informatif. Proses ini menghasilkan pohon keputusan yang dapat dengan mudah diinterpretasikan, sehingga memudahkan dalam memahami faktor-faktor yang mempengaruhi minat masyarakat. C4.5 sangat berguna dalam klasifikasi data dengan struktur yang kompleks dan banyak atribut, serta dapat menangani data dengan missing values dan atribut yang bersifat kategorikal. Di sisi lain, Naive Bayes adalah metode klasifikasi berbasis probabilitas yang mengasumsikan bahwa setiap fitur dalam data independen satu sama lain. Meskipun asumsi independensi ini sering kali tidak sepenuhnya akurat, Naive Bayes masih dapat memberikan hasil klasifikasi yang sangat baik, terutama dalam analisis teks dan media sosial. Metode ini bekerja dengan menghitung probabilitas setiap kelas berdasarkan fitur yang ada dan memilih kelas dengan probabilitas tertinggi sebagai hasil klasifikasi. Keunggulan Naive Bayes terletak pada kemampuannya untuk menangani data besar dan tinggi dimensi dengan efisiensi komputasi yang tinggi, serta kemampuannya untuk memberikan prediksi yang cepat dan akurat dalam analisis minat masyarakat.

#### IV. Hasil dan Pembahasan Langkah-Langkah Pengelolaan Metode Naive Bayes

Analisis Data Untuk tahapan ataupun langkah-langkah pengelolaan data pada metode Naive Bayes, untuk tahapan awalnya yaitu menentukan atribut dan data yang akan digunakan

dan nantinya data tersebut akan diolah. Untuk data yang akan digunakan yaitu sebagai berikut.

**Tabel 1. Data Sampel Penelitian**

Nama Lengkap	Jenis Kelamin	Platform Media Sosial	Kemudahan Penggunaan	Privasi dan Keamanan	Fitur-Fitur	Kualitas Visual
Diah Pramesti Cahyani	Perempuan	Twitter	Sulit	Lemah	Tidak Bagus	Bagus
Farida Azmy Siregar	Perempuan	Instagram	Mudah	Kuat	Bagus	Tidak Bagus
Khairani Karen Azara Siregar	Perempuan	TikTok	Mudah	Lemah	Tidak Bagus	Tidak Bagus
Khairulsyah Siregar	Laki-Laki	Twitter	Mudah	Lemah	Bagus	Bagus
Nurmaulina	Perempuan	Facebook	Sulit	Lemah	Bagus	Tidak Bagus
PikriIhsan Ritonga	Laki-Laki	Facebook	Sulit	Kuat	Bagus	Bagus
Puspita Sari	Perempuan	TikTok	Mudah	Kuat	Bagus	Tidak Bagus
Ratna	Perempuan	Instagram	Sulit	Kuat	Tidak Bagus	Tidak Bagus
Ridho Nufaizal	Laki-Laki	Facebook	Mudah	Kuat	Tidak Bagus	Bagus
Wakiman	Laki-Laki	Facebook	Mudah	Kuat	Bagus	Tidak Bagus

Tabel diatas merupakan data sampel Penelitian yang sudah diperoleh penulis dari hasil kuesioner yang dibagikan kepada para responden yang menjadi sampel penelitian.

#### Data Latih

Untuk data latih merupakan data yang digunakan untuk membantu proses pengolahan data, baik itu secara manual maupun menggunakan sistem aplikasi seperti aplikasi orange dan rapid miner.

Tetapi sebelum perhitungan dimulai, data akan dibagi menjadi beberapa tabel yang dimana setiap table berisi 1 atribut dan value ataupun isi dari setiap atribut. Untuk table nya dapat dilihat pada table dibawah.

**Tabel 2. Atribut Platform Media Sosial**

Atribut	Value	P (Minat)	P (Tidak Minat)
Platform Media Sosial	Facebook	3/6	1/4
	Instagram	1/6	1/4
	Twitter	1/6	1/4
	Tiktok	1/6	1/4
Total		100%	100%

**Tabel 3. Atribut Kemudahan Penggunaan**

Atribut	Value	P (Minat)	P (Tidak Minat)
Kemudahan Penggunaan	Mudah	5/6	1/4
	Sulit	1/6	3/4
Total		100%	100%

**Tabel 4. Atribut Privasi dan Keamanan**

Atribut	Value	P (Minat)	P (Tidak Minat)
Privasidan Keamanan	Kuat	5/6	1/4
	Lemah	1/6	3/4
Total		100%	100%

**Tabel 5. Atribut Fitur-Fitur**

Atribut	Value	P (Minat)	P (Tidak Minat)
Privasidan Keamanan	Kuat	5/6	1/4
	Lemah	1/6	3/4
Total		100%	100%

**Tabel 6. Atribut Kualitas Visual**

Atribut	Value	P (Minat)	P (Tidak Minat)
Fitur-Fitur	Bagus	5/6	1/4
	TidakBagus	1/6	3/4
Total		100%	100%

**Tabel 7. Atribut Kategori**

Atribut	Value	P (Minat)	P (Tidak Minat)
Kualitas Visual	Bagus	3/6	1/4
	TidakBagus	3/6	3/4
Total		100%	100%

Pada tabel diatas merupakan data dari setiap atribut yang terdaat nilai nya. Nilai tersebut diperoleh dari data latih diatas. Jadi untuk melakukan perhitungan manual pada metode Naïve Bayes, harus melakukan pemisahan ataupun pembagian tabel

berdasarkan setiap atribut yang ada.

**Perhitungan**

Untuk perhitungan pada metode Naive Bayes menggunakan rumus sebagai berikut.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

P(A|B) = Probabilitas A bersyarat yang diberikan oleh B

P(B|A) = Probabilitas B bersyarat yang diberikan oleh A

P(A) = Probabilitas kejadian A

P(B) = Probabilitas kejadian B

Untuk perhitungan yang akan dilakukan, pertama penulis akan menghitung data. Adapun perhitungannya sebagai berikut.

$$\begin{aligned}
 P(\text{Kategori}) &= P(\text{Platform Media Sosial}|\text{Twitter}) \times P(\text{Kemudahan Penggunaan}|\text{Sulit}) \times \\
 & P(\text{Privasi dan Keamanan}|\text{Lemah}) \times P(\text{Fitur-Fitur}|\text{Tidak Bagus}) \times P(\text{Kualitas Visual}|\text{Bagus}) \times P(\text{Kategori}|\text{Minat}) \\
 P(\text{Kategori}) &= \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{3}{6} \times \frac{6}{10} \times P(\text{Kategori}|\text{Minat}) \\
 &= \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{3}{6} \times \frac{6}{10} \times P(\text{Kategori}|\text{Minat}) \\
 &= 0,000231481 (\text{Nilai Minat}) \\
 P(\text{Kategori}) &= \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{4}{10} \times P(\text{Kategori}|\text{Tidak Minat}) \\
 &= \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{4}{10} \times P(\text{Kategori}|\text{Tidak Minat}) \\
 &= 0,010546875 (\text{Nilai Tidak Minat})
 \end{aligned}$$

Jadi dari hasil kedua perhitungan diatas antara nilai minat dan tidak minat hasilnya lebih besar nilai tidak minat, maka hasil perhitungan tersebut menyatakan tidak minat, hal ini karena nilai tidak minat lebih besar daripada nilai minat. Jadi data Diah Pramesti Cahyani masuk pada kategori tidak minat pada media sosial. Untuk menghitung data selanjutnya sama saja seperti



perhitungan diatas. Maka hasil untuk semua data yang ada yaitu sebanyak 10 data penulis buat dalam bentuk tabel. Untuk data tabelnya seperti berikut.

**Tabel 8. Perhitungan didalam Tabel**

Nama Lengkap		Pengerjaan	Nilai	Kategori
Diah Pramesti Cahyani	P (Kategori Minat)	$= (1/8) * (1/8) * (1/8) * (1/8) * (5/8) * (8/10)$	0.000251481	Tidak Minat
	P (Kategori Tidak Minat)	$= (1/4) * (5/4) * (5/4) * (5/4) * (1/4) * (4/10)$	0.010240813	
Fanda Army Siregar	P (Kategori Minat)	$= (1/8) * (5/8) * (1/8) * (5/8) * (5/8) * (8/10)$	0.00281051	Minat
	P (Kategori Tidak Minat)	$= (1/4) * (1/4) * (1/4) * (1/4) * (5/4) * (4/10)$	0.00111813	
Khairani Karen Anas Siregar	P (Kategori Minat)	$= (1/8) * (5/8) * (1/8) * (1/8) * (5/8) * (8/10)$	0.001151401	Tidak Minat
	P (Kategori Tidak Minat)	$= (1/4) * (1/4) * (5/4) * (5/4) * (5/4) * (4/10)$	0.010240813	
Khairuliyah Siregar	P (Kategori Minat)	$= (1/8) * (5/8) * (1/8) * (5/8) * (5/8) * (8/10)$	0.00281051	Minat
	P (Kategori Tidak Minat)	$= (1/4) * (1/4) * (5/4) * (1/4) * (1/4) * (4/10)$	0.00111813	
Nurmalina	P (Kategori Minat)	$= (5/8) * (1/8) * (1/8) * (5/8) * (5/8) * (8/10)$	0.00342222	Tidak Minat
	P (Kategori Tidak Minat)	$= (1/4) * (5/4) * (1/4) * (1/4) * (5/4) * (4/10)$	0.003512025	
Pikri Ihsan Ritonga	P (Kategori Minat)	$= (5/8) * (1/8) * (5/8) * (5/8) * (5/8) * (8/10)$	0.01561111	Minat
	P (Kategori Tidak Minat)	$= (1/4) * (5/4) * (1/4) * (1/4) * (1/4) * (4/10)$	0.00111813	
Puspita Sari	P (Kategori Minat)	$= (1/8) * (5/8) * (5/8) * (5/8) * (5/8) * (8/10)$	0.01561111	Minat
	P (Kategori Tidak Minat)	$= (1/4) * (1/4) * (1/4) * (1/4) * (5/4) * (4/10)$	0.00111813	
Ratna	P (Kategori Minat)	$= (1/8) * (1/8) * (5/8) * (1/8) * (5/8) * (8/10)$	0.001151401	Tidak Minat
	P (Kategori Tidak Minat)	$= (1/4) * (5/4) * (1/4) * (5/4) * (5/4) * (4/10)$	0.010240813	
Ridho Nufazal	P (Kategori Minat)	$= (5/8) * (5/8) * (5/8) * (1/8) * (5/8) * (8/10)$	0.01561111	Minat
	P (Kategori Tidak Minat)	$= (1/4) * (1/4) * (1/4) * (5/4) * (1/4) * (4/10)$	0.00111813	
Wakiman	P (Kategori Minat)	$= (5/8) * (5/8) * (5/8) * (5/8) * (5/8) * (8/10)$	0.00880250	Minat
	P (Kategori Tidak Minat)	$= (1/4) * (1/4) * (1/4) * (1/4) * (5/4) * (4/10)$	0.00111813	

**Langkah-Langkah Algoritma C4.5 Analisis Data**

Untuk proses pengelolaan algoritma c4.5, tahapan pertama adalah menentukan data yang akan dihitung. Untuk data yang dihitung sudah

terdapat kategori dari setiap data. Hal ini karena pada metode ini untuk pengerjaan manualnya untuk menentukan bentuk pohon keputusan dari algoritma C4.5. Untuk data sampelnya yaitu sebagai berikut.

**Tabel 9. Data Sampel Algoritma C4.5**

Nama Lengkap	Jenis Kelamin	Platform Media Sosial	Kemudahan Penggunaan	Privasi dan Keamanan	Fitur-Fitur	Kualitas Visual	Kategori
Diah Pramesti Cahyani	Perempuan	Twitter	Sulit	Lemah	Tidak Bagus	Bagus	Tidak Minat
Fanda Army Siregar	Perempuan	Instagram	Mudah	Kuat	Bagus	Tidak Bagus	Minat
Khairani Karen Anas Siregar	Perempuan	TikTok	Mudah	Lemah	Tidak Bagus	Tidak Bagus	Minat
Khairuliyah Siregar	Laki-Laki	Twitter	Mudah	Lemah	Bagus	Bagus	Tidak Minat
Nurmalina	Perempuan	Facebook	Sulit	Lemah	Bagus	Tidak Bagus	Tidak Minat
Pikri Ihsan Ritonga	Laki-Laki	Facebook	Sulit	Kuat	Bagus	Bagus	Minat
Puspita Sari	Perempuan	TikTok	Mudah	Kuat	Bagus	Tidak Bagus	Minat
Ratna	Perempuan	Instagram	Sulit	Kuat	Tidak Bagus	Tidak Bagus	Tidak Minat
Ridho Nufazal	Laki-Laki	Facebook	Mudah	Kuat	Tidak Bagus	Bagus	Minat
Wakiman	Laki-Laki	Facebook	Mudah	Kuat	Bagus	Tidak Bagus	Minat

Pada tabel diatas merupakan data yang akan digunakan sebagai data sampel penelitian untuk perhitungan manual algoritma c4.5. Pada proses pengerjaan algoritma c4.5 menggunakan rumus pada metode decision tree, rumus yang akan digunakan yaitu rumus entropy untuk menghitung nilai setiap atribut dan rumus gain yang digunakan untuk menghitung nilai keseluruhan dari atribut yang ada pada data sampel. Untuk rumusnya dapat dilihat sebagai berikut.

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i$$

Keterangan:

S = Himpunan kasus

n = Jumlah Partisi S

P<sub>i</sub> = Proporsi S<sub>i</sub> terhadap S

$$\log_2(x) = \frac{\ln(x)}{\ln(2)}$$

$$\log_2 P_i = \frac{\ln(P_i)}{\ln(2)}$$

$$Gain(S, A) = entropy(S) - \sum_{i=1}^{|S_i|} * Entropy(S_i)$$

Keterangan:

S = Himpunan kasus A = Fitur

n = Jumlah Partisi atribut A |S<sub>i</sub>| = Proporsi S<sub>i</sub>

terhadap S

|S<sub>i</sub>| = Jumlah kasus dalam S

Untuk memulia perhitungan, terlebih dahulu penulis menghitung nilai entropy dari setiap partisi yang ada pada setiap atribut. Kemudian setelah setiap partisi dihitung maka kemudian penulis akan menghitung gain dari setiap atribut. Sebelumnya data sampel akan dipecah dan dibagi menjadi beberapa table sebanyak atribut yang digunakan. Tabel tersebut yaitu sebagai berikut.

**Tabel 10. Platform Media Sosial**

Atribut	Partisi	Kasus (S)	Tidak Minat (S1)	Minat (S2)
Platform Media Sosial	Twitter	2	1	1
	Instagram	2	1	1
	Facebook	4	1	3
	TikTok	2	1	1

**Tabel 11. Kemudahan Penggunaan**

Atribut	Partisi	Kasus (S)	Tidak Minat (S1)	Minat (S2)
Kemudahan Penggunaan	Mudah	6	1	5
	Sulit	4	3	1

**Tabel 12. Privasi dan Keamanan**

Atribut	Partisi	Kasus (S)	Tidak Minat (S1)	Minat (S2)
Privasi dan Keamanan	Kuat	6	1	5
	Lemah	4	3	1

**Tabel 13. Fitur-Fitur**

Atribut	Partisi	Kasus (S)	Tidak Minat (S1)	Minat (S2)
Fitur-Fitur	Bagus	6	1	5
	Tidak Bagus	4	3	1

**Tabel 14. Kualitas Visual**

Atribut	Partisi	Kasus (S)	Tidak Minat (S1)	Minat (S2)
Kualitas Visual	Bagus	4	3	1
	Tidak Bagus	6	3	3

Pada beberapa tabel diatas merupakan data sampel yang sudah dibagi menjadi beberapa table. Hal ini dilakukan agar data dapat dihitung.

**Menghitung Entropy Total**

$$\begin{aligned}
 \text{Total data (S)} &= 10 \\
 \text{Jumlah Kasus "Tidak Minat"} &= 4 \\
 \text{Jumlah Kasus "Minat"} &= 6 \\
 \text{Entropy (Total)} &= \left(-\frac{4}{10} \times \text{Log}_2 \left(\frac{4}{10}\right)\right) + \left(-\frac{6}{10} \times \text{Log}_2 \left(\frac{6}{10}\right)\right) \\
 &= \left(-\frac{4}{10} \times \left(\frac{\ln\left(\frac{4}{10}\right)}{\ln(2)}\right)\right) + \left(-\frac{6}{10} \times \left(\frac{\ln\left(\frac{6}{10}\right)}{\ln(2)}\right)\right) \\
 &= \left(-0,4 \times \left(-\frac{0,91629}{0,69314}\right)\right) + \left(-0,6 \times \left(-\frac{0,57082}{0,69314}\right)\right) \\
 &= (-0,4) \times (-1,32194) + (-0,6) \times (-0,73696) \\
 &= 0,52877 + 0,44217 \\
 &= 0,97094
 \end{aligned}$$

**Menghitung Entropy Platform Media Sosial "Twitter"**

$$\begin{aligned}
 \text{Entropy (Total)} &= \left(-\frac{1}{2} \times \text{Log}_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \text{Log}_2 \left(\frac{1}{2}\right)\right) \\
 &= \left(-\frac{1}{2} \times \left(\frac{\ln\left(\frac{1}{2}\right)}{\ln(2)}\right)\right) + \left(-\frac{1}{2} \times \left(\frac{\ln\left(\frac{1}{2}\right)}{\ln(2)}\right)\right) \\
 &= \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) + \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) \\
 &= (-0,5) \times (-1) + (-0,5) \times (-1) \\
 &= 0,5 + 0,5 \\
 &= 0,1
 \end{aligned}$$

**Menghitung Entropy Platform Media Sosial "Instagram"**

$$\begin{aligned}
 \text{Entropy (Total)} &= \left(-\frac{1}{2} \times \text{Log}_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \text{Log}_2 \left(\frac{1}{2}\right)\right) \\
 &= \left(-\frac{1}{2} \times \left(\frac{\ln\left(\frac{1}{2}\right)}{\ln(2)}\right)\right) + \left(-\frac{1}{2} \times \left(\frac{\ln\left(\frac{1}{2}\right)}{\ln(2)}\right)\right) \\
 &= \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) + \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) \\
 &= (-0,5) \times (-1) + (-0,5) \times (-1) \\
 &= 0,5 + 0,5 \\
 &= 0,1
 \end{aligned}$$

Menghitung Entropy Platform Media Sosial “TikTok”

$$\begin{aligned} \text{Entropy (Total)} &= \left(-\frac{1}{2} \times \text{Log}_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \text{Log}_2 \left(\frac{1}{2}\right)\right) \\ &= \left(-\frac{1}{2} \times \frac{\ln\left(\frac{1}{2}\right)}{\ln(2)}\right) + \left(-\frac{1}{2} \times \frac{\ln\left(\frac{1}{2}\right)}{\ln(2)}\right) \\ &= \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) + \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) \\ &= (-0,5) \times (-1) + (-0,5) \times (-1) \\ &= 0,5 + 0,5 \\ &= 0,1 \end{aligned}$$

Menghitung Gain Platform Media Sosial

$$\begin{aligned} \text{Entropy (Total)} &= 0,97094 - \left(\frac{2}{10} \times 1\right) + \left(\frac{2}{10} \times 1\right) + \left(\frac{4}{10} \times 0,81127\right) + \left(\frac{2}{10} \times 1\right) \\ &= 0,97094 \times (0,2) + (0,2) \times (0,3245) + (0,2) \\ &= 0,97094 + 0,9245 \\ &= 0,04644 \end{aligned}$$

Menghitung Entropy Kemudahan Penggunaan “Mudah”

$$\begin{aligned} &= 0,69314 \quad 0,69314 \\ &= (0,16666) \times (-2,58497) + (-0,83333) \times (-0,26303) \\ &= 0,43081 + 0,21919 \\ &= 0,65 \end{aligned}$$

Menghitung Entropy Kemudahan Penggunaan “Sulit”

$$\begin{aligned} \text{Entropy (Total)} &= \left(-\frac{3}{4} \times \text{Log}_2 \left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} \times \text{Log}_2 \left(\frac{1}{4}\right)\right) \\ &= \left(-\frac{3}{4} \times \frac{\ln\left(\frac{3}{4}\right)}{\ln(2)}\right) + \left(-\frac{1}{4} \times \frac{\ln\left(\frac{1}{4}\right)}{\ln(2)}\right) \\ &= \left(-0,75 \times \left(-\frac{1,28768}{0,69314}\right)\right) + \left(-0,25 \times \left(-\frac{1,38629}{0,69314}\right)\right) \\ &= (0,75) \times (-0,41503) + (-0,25) \times (-2,00001) \\ &= 0,31127 + 0,5 \\ &= 0,81127 \end{aligned}$$

Menghitung Gain Kemudahan Penggunaan

$$\begin{aligned} \text{Entropy (Total)} &= 0,97094 - \left(\frac{6}{10} \times 0,65\right) + \left(\frac{4}{10} \times 0,81127\right) \\ &= 0,97094 \times (0,39) + (0,3245) \\ &= 0,97094 + 0,7145 \\ &= 0,25644 \end{aligned}$$

Menghitung Entropy dan Gain Kualitas Visual Menghitung Entropy Kualitas Visual “Bagus”

$$\begin{aligned} &\frac{3}{4} \quad \left(\frac{3}{4}\right) \quad \frac{1}{4} \quad \left(\frac{1}{4}\right) \\ &\frac{3}{4} \quad \left(\frac{\ln\left(\frac{3}{4}\right)}{\ln(2)}\right) \quad \frac{1}{4} \quad \left(\frac{\ln\left(\frac{1}{4}\right)}{\ln(2)}\right) \\ &= \left(-0,75 \times \left(-\frac{1,28768}{0,69314}\right)\right) + \left(-0,25 \times \left(-\frac{1,38629}{0,69314}\right)\right) \\ &= (0,75) \times (-0,41503) + (-0,25) \times (-2,00001) \\ &= 0,31127 + 0,5 \\ &= 0,81127 \end{aligned}$$

Menghitung Entropy Kualitas Visual “Tidak Bagus”

$$\begin{aligned} \text{Entropy (Total)} &= \left(-\frac{3}{6} \times \text{Log}_2 \left(\frac{3}{6}\right)\right) + \left(-\frac{3}{6} \times \text{Log}_2 \left(\frac{3}{6}\right)\right) \\ &= \left(-\frac{3}{6} \times \frac{\ln\left(\frac{3}{6}\right)}{\ln(2)}\right) + \left(-\frac{3}{6} \times \frac{\ln\left(\frac{3}{6}\right)}{\ln(2)}\right) \\ &= \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) + \left(-0,25 \times \left(-\frac{0,69314}{0,69314}\right)\right) \\ &= (0,5) \times (-0,1) + (-0,5) \times (-1) \\ &= 0,5 + 0,5 \\ &= 1 \end{aligned}$$

Menghitung Gain Fitur-Fitur

$$\begin{aligned} \text{Entropy (Total)} &= 0,97094 - \left(\frac{6}{10} \times 0,65\right) + \left(\frac{4}{10} \times 0,81127\right) \\ &= 0,97094 \times (0,39) + (0,3245) \\ &= 0,97094 + 0,7145 \\ &= 0,25644 \end{aligned}$$

Menghitung Entropy dan Gain Kualitas Visual Menghitung Entropy Kualitas Visual “Bagus”

$$\frac{3}{4} \left(\frac{3}{4}\right) \frac{1}{4} \left(\frac{1}{4}\right)$$

$$\frac{3}{4} \left(\frac{\ln(\frac{3}{4})}{\ln(2)}\right) \frac{1}{4} \left(\frac{\ln(\frac{1}{4})}{\ln(2)}\right)$$

$$(-0,75 \times (-\frac{1,28768}{0,69314})) + (-0,25 \times (-\frac{1,38629}{0,69314}))$$

$$= (0,75) \times (-0,41503) + (-0,25) \times (-2,00001)$$

$$= 0,31127 + 0,5$$

$$= 0,81127$$

Menghitung Entropy Kualitas Visual  
“Tidak Bagus”

$$\text{Entropy (Total)} = \left(-\frac{3}{6} \times \text{Log}_2\left(\frac{3}{6}\right)\right) + \left(-\frac{3}{6} \times \text{Log}_2\left(\frac{3}{6}\right)\right)$$

$$= \left(-\frac{3}{6} \times \frac{\ln(\frac{3}{6})}{\ln(2)}\right) + \left(-\frac{3}{6} \times \frac{\ln(\frac{3}{6})}{\ln(2)}\right)$$

$$= \left(-0,5 \times \left(-\frac{0,69314}{0,69314}\right)\right) + \left(-0,25 \times \left(-\frac{0,69314}{0,69314}\right)\right)$$

$$= (0,5) \times (-0,1) + (-0,5) \times (-1)$$

$$= 0,5 + 0,5$$

$$= 1$$

Menghitung Gain Kualitas Visual

$$\text{Entropy (Total)} = 0,97094 - \left(\frac{6}{10} \times 1\right) + \left(\frac{4}{10} \times 0,81127\right)$$

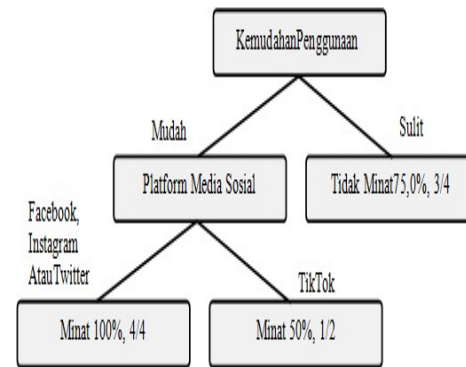
$$= 0,97094 \times (0,6) + (0,3245)$$

$$= 0,97094 + 0,9245$$

$$= 0,04644$$

Tabel 15. Menghitung Gain Kualitas Visual

Atribut	Partisi	Kasus (S)	Tidak Minat (S1)	Minat (S2)	Entropy	Gain
Total					0,97094	
Platform Media Sosial						0,04644
	Twitter	2	1	1	1	
	Instagram	2	1	1	1	
	Facebook	4	1	3	0,81127	
	TikTok	2	1	1	1	
Kemudahan Penggunaan						0,25644
	Mudah	6	1	5	0,65	
	Sulit	4	3	1	0,81127	
Privasi dan Keamanan						0,25644
	Kuat	6	1	5	0,65	
	Lemah	4	3	1	0,81127	
Fitur-Fitur						0,25644
	Bagus	6	1	5	0,65	
	Tidak Bagus	4	3	1	0,81127	
Kualitas Visual						0,04644
	Bagus	4	3	1	0,81127	
	Tidak Bagus	6	3	3	1	



Gambar 5. Hasil Entropy

Hasil dari perhitungan yang sudah dilakukan yaitu memperoleh tree wiewer. Hasil tersebut diperoleh dari hasil entropy.

## V. Kesimpulan

Analisis minat masyarakat menggunakan media sosial dalam data mining dapat dilakukan dengan efektif menggunakan metode algoritma C4.5 dan Naive Bayes. Algoritma C4.5 menghasilkan pohon keputusan yang memudahkan identifikasi faktor-faktor utama yang mempengaruhi minat, sementara Naive Bayes memberikan klasifikasi berbasis probabilitas yang efisien dan akurat. Kedua metode ini,



ketika diterapkan pada data media sosial, mampu melakukan klasifikasi dengan sangat baik, menghasilkan hasil yang sangat akurat dan informatif. Hal ini menunjukkan kemampuan kedua metode dalam menangani data yang kompleks dan tinggi dimensi, serta dalam memahami pola minat masyarakat secara mendalam.

## VI. Daftar Pustaka

- A. H. M. Alaidi, I. A. Aljazaery, and S. H. Abbood, "Dark Web Illegal Activities Crawling and Classifying Using Data Mining Techniques," vol. 1, pp. 122–139.
- A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- Awang Hendrianto Pratomo, W. Kaswidjanti, and S. Mu'arifah, "Implementasi Algoritma Region of Interest (ROI) Untuk Meningkatkan Performa Algoritma Deteksi Dan Klasifikasi Kendaraan," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 155–162, 2020, doi: 10.25126/jtiik.202071718.
- D.A.C, N. Bayes, and D.A.N. Svm, "PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA," vol. 13, no. 1, pp. 16–25, 2019
- F. F. Dwi Imaniawan, R. Wijianto, and J. D. Mulyanto, "Penerapan Search Engine Optimization Untuk Meningkatkan Peringkat Website Pada Search Engine Results Page Google (Studi Kasus: www.superbengkel.co.id)," *EVOLUSIJ. Sains dan Manaj.*, vol. 8, no. 2, 2020, doi: 10.31294/evolusi.v8i2.9172.
- H. B. Jatmiko, N. Tedi Kurniadi, and D. Maulana, "Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta," *J. Autom. Comput. Inf. Syst.*, vol. 2, no. 1, pp. 22–30, 2022, doi: 10.47134/jacis.v2i1.35.
- Hermantoro, A. P. Suryotomo, A. I. Uktoro, and R. A. Renjani, "Unmanned Aerial Vehicle Application for Plantation Mapping and Automatic Oil Palm Trees Counting on Oil Palm Plantation Management," in *International Conference on the Role of Agricultural Engineering for Sustainable Agriculture Production*, 2016, pp. 47–50.
- I. P. Ninditama, I. P. Ninditama, W. Cholil, M. Akbar, and D. Antoni, "Klasifikasi Keluarga Sejahtera Study Kasus: Kecamatan Kota Palembang," vol. 15, no. 2, pp. 37–49, 2020.
- K. A. Mahasiswa, R. Rachmatika, and A. Bisri, "Perbandingan Model Klasifikasi untuk Evaluasi," vol. 6, no. 3, pp. 417–422, 2020.
- M.A. Afifi, T.M. Ghazal, M.A.M. Afifi, and D. Kalra, "Data Mining and Exploration: A Comparison Study among Data Mining Techniques on Iris Data Set," *Talent Dev. Excell.*, vol. 12, no. 1,

- pp.3854–3861,2020,[Online].Available:ht  
tp://www.iratde.com
- M.Adriansa,L.Yulianti,L.Elfianty,andU.  
D.Bengkulu,“AnalisisKepuasanP  
elanggan Menggunakan  
Algoritma,” vol. 07, no. 21, pp.  
115–121, 2022.
- N. Dengan *et al.*, “PENERAPAN  
DATA MINING UNTUK  
MEMREDIKSI  
TRANSAKSI,” vol. 2, no. 1, pp.  
23–28, 2021.
- R. Aztin and K. A. Musodo, “Penerapan  
Text Mining Dengan Algoritma  
Naïve Bayes Untuk  
Mengklasifikasikan  
SentimenRakyatTerhadapMinya  
kGorengSubsidiPemerintah,”  
*Semin. Nas. Mhs. Fak. ...*, no.  
September, pp. 645–652, 2022,  
[Online].Available:  
<http://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/347%0Ahttp://senafiti.budiluhur.ac.id/index.php/senafiti/article/download/347/76>
- R. Forest and N. Bayes, “Perbandingan  
Akurasi , Recall , dan Presisi  
Klasifikasi pada Algoritma,”  
vol. 5, no. April, pp. 640–651,  
2021, doi:  
10.30865/mib.v5i2.2937.
- R. Ratra and P. Gulia, “Experimental  
evaluation of open source data  
mining tools  
(WEKAandorange),”*Int.J.Eng.T  
rendsTechnol.*,vol.68,no.8,pp.30  
–35,2020,doi:  
10.14445/22315381/IJETT-  
V68I8P206S
- R.T.Aldisaand P.Maulana,“Analisis  
Sentimen Opini Masyarakat  
TerhadapVaksinasi Booster  
COVID- 19 Dengan  
Perbandingan Metode Naive  
Bayes , Decision Tree dan  
SVM,” vol. 4, no. 1, pp. 106–  
109, 2022, doi:  
10.47065/bits.v4i1.1581.
- S. P. Tahalea and A. SN, “Central Actor  
Identification of Crime Group  
using Semantic Social Network  
Analysis,” *Indones. J. Inf. Syst.*,  
vol. 2, no. 1, p. 24, 2019, doi:  
10.24002/ijis.v2i1.2354.
- S. P. Tahalea, “Identifikasi Peran Hero  
DOTA2 Menggunakan Social  
Network Analysis,”  
*TEKNOMATIKA*, vol. 12, no. 2,  
pp. 81–86, 2020.
- S. Saifullah and V. A. Permadi,  
“Comparison of Egg Fertility  
Identification based on  
GLCMFeatureExtractionusingB  
ackpropagationandK-  
meansClusteringAlgorithms,” in  
*Proceeding - 2019 5th  
International Conference on  
Science in Information  
Technology: Embracing  
Industry 4.0: Towards  
Innovation in Cyber Physical  
System*,  
*ICSITech2019*,InstituteofElectric  
alandElectronicsEngineersInc.,O  
ct.2019,pp.140–145. doi:  
10.1109/ICSITech46713.2019.89  
87496.
- Y. Indah Lestari and S. Defit, “Jurnal  
Informatika Ekonomi Bisnis  
Prediksi Tingkat Kepuasan  
Pelayanan Online Menggunakan  
Metode Algoritma C.45,” vol. 3,  
pp. 148– 154, 2021, doi:  
10.37034/infeb.v3i3.104.