

# SENTIMEN ANALISIS PENGAJUAN AMICUS CURIAE KE MAHKAMAH KONSTITUSI MENGGUNAKAN METODE NAIVE BAYES

Muhammad Daffa Aditya<sup>1\*</sup>, Utomo Budiyanto<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

E-mail: <sup>1\*</sup>2011501380@student.budiluhur.ac.id, <sup>2</sup>utomo.budiyanto@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak-***Amicus Curiae* adalah pihak yang memberikan pandangan atau opini terkait suatu kasus kepada pengadilan, meskipun bukan merupakan pihak langsung dalam kasus tersebut. Kasus umum pada tahun 2024 di Indonesia menjadi salah satu isu besar yang mendapatkan perhatian luas dari masyarakat, termasuk di media sosial. Pengajuan *Amicus Curiae* dalam konteks ini menarik perhatian banyak pihak karena mencerminkan bagaimana masyarakat dan berbagai kelompok memberikan pandangan hukum dan sosial mereka terhadap proses pemilu. Analisis terhadap ulasan pengguna media sosial terkait pengajuan *Amicus Curiae* ke Mahkamah Konstitusi dapat memberikan gambaran yang lebih jelas mengenai opini publik dan sentimen yang berkembang. Penelitian ini bertujuan untuk menganalisis sentimen dan opini pengguna media sosial *X* terkait dengan pengajuan *Amicus Curiae* ke Mahkamah Konstitusi mengenai kasus umum yang ada di Indonesia tahun 2024. Hasil dari penelitian menggunakan metode *Naive Bayes* untuk sentimen otomatis yaitu mendapatkan nilai akurasi sebesar 92,2%, dengan memiliki nilai presisi sebesar 0,92, dan nilai *recall* sebesar 1,0 yang di dominasi oleh kelas sentimen netral. Sedangkan dari data sentimen pakar mendapatkan nilai akurasi sebesar 99%, dengan memiliki nilai presisi sebesar 0,99, dan nilai *recall* sebesar 1.0 yang di dominasi oleh kelas sentimen netral. Dari data *tweet* yang sudah melalui proses pelabelan data, dapat diketahui bahwa pendapat pengguna media sosial *X* memiliki pandangan netral terhadap pengajuan *Amicus Curiae* ke Mahkamah Konstitusi. Selain itu, hasil sentimen analisis juga dapat menjadi dasar bagi pihak terkait untuk mengambil kebijakan atau langkah-langkah strategis yang sesuai dengan aspirasi masyarakat.

**Kata Kunci:** Sentimen analisis, *Amicus Curiae*, *Naive Bayes*, Media Sosial, Mahkamah Konstitusi.

## SENTIMENT ANALYSIS OF AMICUS CURIAE SUBMISSION TO THE CONSTITUTIONAL COURT USING THE NAIVE BAYES METHOD

**Abstract-***Amicus Curiae* is a party that provides views or opinions regarding a case to the court, although not a direct party in the case. General cases in 2024 in Indonesia are one of the major issues that have received widespread attention from the public, including on social media. The submission of *Amicus Curiae* in this context has attracted the attention of many parties because it reflects how the public and various groups provide their legal and social views on the election process. Analysis of social media user reviews regarding the submission of *Amicus Curiae* to the Constitutional Court can provide a clearer picture of public opinion and developing sentiments. This study aims to analyze the sentiments and opinions of social media users *X* related to the submission of *Amicus Curiae* to the Constitutional Court regarding general cases in Indonesia in 2024. The results of the study using the *Naive Bayes* method for automatic sentiment obtained an accuracy value of 92.2%, with a precision value of 0.92, and a recall value of 1.0 which is dominated by the neutral sentiment class. Meanwhile, from the expert sentiment data, the accuracy value is 99%, with a precision value of 0.99, and a recall value of 1.0 which is dominated by the neutral sentiment class. From the *tweet* data that has gone through the data labeling process, it can be seen that the opinions of social media users *X* have a neutral view of the submission of *Amicus Curiae* to the Constitutional Court. In addition, the results of sentiment analysis can also be a basis for related parties to take policies or strategic steps that are in accordance with the aspirations of the community.

**Keywords:** Sentiment analysis, *Amicus Curiae*, *Naive Bayes*, Social Media, Constitutional Court.

### 1. PENDAHULUAN

Dalam era digital ini, media sosial telah menjadi platform utama bagi masyarakat untuk mengemukakan pendapat, berbagi informasi, dan berinteraksi dengan berbagai isu yang berkembang. Media sosial *Twitter* yang sudah berubah nama menjadi *X* merupakan media sosial yang digunakan oleh masyarakat luas sebagai alat berbagi informasi pada pengguna atau linimasa secara *real-time* dengan cara menambahkan komentar terkait pengalaman dan pemikiran mereka. Hal ini tidak hanya mempengaruhi kehidupan sosial, tetapi juga memiliki dampak signifikan dalam ranah politik dan hukum.

Salah satu aspek penting dari interaksi antara masyarakat dan hukum adalah konsep *Amicus Curiae*. *Amicus Curiae*, yang dalam bahasa Latin berarti "sahabat pengadilan," adalah pihak atau kelompok yang memberikan pandangan atau pendapat kepada pengadilan mengenai suatu perkara, meskipun mereka bukan merupakan pihak yang langsung terkait dalam kasus tersebut.

Kasus umum pada tahun 2024 di Indonesia menjadi salah satu isu besar yang mendapatkan perhatian luas dari masyarakat, termasuk di media sosial. Pengajuan *Amicus Curiae* dalam konteks ini menarik perhatian banyak pihak karena mencerminkan bagaimana masyarakat dan berbagai kelompok memberikan pandangan hukum dan sosial mereka terhadap proses pemilu. Analisis terhadap ulasan pengguna media sosial terkait pengajuan *Amicus Curiae* ke Mahkamah Konstitusi dapat memberikan gambaran yang lebih jelas mengenai opini publik dan sentimen yang berkembang. Hal berikut yang membuat penulis tertarik untuk melakukan penelitian dalam sentimen analisis pengguna media sosial X berjudul “SENTIMEN ANALISIS PENGAJUAN AMICUS CURIAE KE MAHKAMAH KONSTITUSI MENGGUNAKAN METODE NAIVE BAYES”.

Terdapat penelitian sebelumnya yang memiliki studi kasus dan metode yang sama, yang dibuat oleh Tommy Gunawan dkk pada Juni 2024 yang berjudul “ANALISIS SENTIMEN PEMBACAAN PUTUSAN MK TERKAIT PILPRES 2024 PADA APLIKASI X DENGAN METODE NAIVE BAYES CLASSIFIER” yang membahas reaksi publik terhadap keputusan MK sangat beragam, mencerminkan ketidakpuasan dan penerimaan di kalangan masyarakat. Dengan akurasi model mencapai 87%. Pada klasifikasi kategori “*Negative*”, model mencapai presisi sebesar 80%, *recall* sebesar 85%, dan *F1-score* sebesar 82%, menunjukkan keseimbangan yang baik dalam performa klasifikasi. Untuk kategori “*Positive*”, model menunjukkan *presisi* sebesar 88%, *recall* sebesar 84%, dan *F1-score* sebesar 86%, mengindikasikan akurasi dan konsistensi tinggi dalam identifikasi data berlabel positif [1].

Kemudian untuk penelitian lainnya yang memiliki studi kasus yang berbeda, tetapi memiliki metode yang sama oleh Windu Gata dan Achmad Bayhaqy pada Agustus 2020 yang berjudul “ANALYSIS SENTIMENT ABOUT ISLAMOPHOBIA WHEN CHRISTCHURCH ATTACK ON SOCIAL MEDIA” yang membahas reaksi pengguna *Twitter* mengenai serangan teror mengerikan yang terjadi di masjid *Christchurch* di Selandia Baru. Dengan hasil akurasi sebesar 91.390% untuk penggunaan metode SVM dan SMOTE daripada hasil sebelumnya sebesar 73.128% dengan hanya menggunakan metode *Naive Bayes* [2].

Oleh karena itu, penelitian ini bertujuan untuk memberikan pemahaman yang lebih dalam mengenai sikap dan pandangan masyarakat terhadap pengajuan *Amicus Curiae* dalam konteks kasus umum di Indonesia pada tahun 2024. Untuk metode penelitian ini, menggunakan algoritma *Naive Bayes* untuk mengklasifikasikan ulasan sebagai positif, negatif atau netral, dan membantu dalam memberikan akurasi yang lebih tinggi.

## 2. METODE PENELITIAN

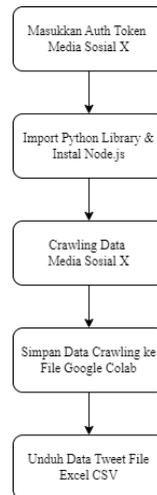
### 2.1 Sentimen Analisis

Sentimen analisis atau penambangan opini itu merujuk kepada penerapan pemrosesan bahasa alami, *linguistic* komputasi dan analisis teks untuk mengidentifikasi dan mengklasifikasikan pendapat subjektif dalam sumber materi. Sentimen analisis bertujuan untuk menentukan sikap terhadap beberapa topik maupun polaritas kontekstual keseluruhan dari sebuah dokumen [3]. Sentimen analisis mengklasifikasikan sentimen berdasarkan polaritas teks dalam sebuah frasa, sehingga dapat menentukan sebagai kelas sentimen positif, negatif, atau netral [4].

### 2.2 Pengumpulan Data

Pengumpulan data atau *Crawling* adalah suatu teknik yang digunakan untuk mengumpulkan informasi yang ada dalam *web*. [5]. Data yang digunakan dalam melakukan penelitian ini berasal dari media sosial X dan terdiri dari *tweet* yang dikumpulkan antara tanggal 19 April 2024 hingga 27 Mei 2024. Jumlah total data yang berhasil diperoleh adalah 509 data *tweet*.

Dalam Gambar 1, dilakukan proses penggunaan proses *crawling data* yang memerlukan beberapa langkah. Untuk langkah-langkahnya yaitu mengunduh aplikasi *Node.js* sebagai *library* untuk mengumpulkan data media sosial X dan menghubungkannya di *Google Colab* dengan memasukan *twitter\_auth\_token* dari akun media sosial X. Kemudian, diperlukannya untuk meng-*import* data-data yang mendukung bahasa pemrograman *Python* seperti *pandas* dan juga menggunakan *Node.js* sebagai penggunaan *crawling data* model *twitter-harvest*. Selanjutnya, untuk men-*crawling* data yang dapat disimpan melalui input filename dalam bentuk *csv*, untuk bagian mencari datanya dalam bentuk input *search\_keyword*, dapat membatasi jumlah data dalam input *limit*, dan diakhiri dengan *command* khusus untuk *twitter-harvest*. Data *tweet* yang berhasil dikumpulkan akan disimpan dalam *files* dari *Google Colab*-nya, lalu mengunduhnya lagi menjadi File *Excel* model *csv*.



**Gambar 1.** Pengumpulan Data

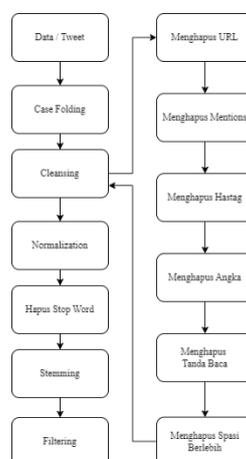
### 2.3 Text Mining

*Text mining* dalam definisi *Data mining* menurut Gartner merupakan proses menemukan korelasi, pola, dan tren baru yang bermakna dengan menyaring sejumlah besar data yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika [6]. Dan menurut Sukaya dan Biruntha dalam Salloum et.al, *text mining* ibarat sebuah sistem kecerdasan yang mengekstraksi kata atau kalimat yang tepat dari kata yang tidak tepat dan kemudian mengubah kata tersebut menjadi saran tertentu. Penambangan teks pada dasarnya adalah bidang baru yang memiliki tujuan utama pemulihan data, pembelajaran mesin, penambangan informasi, dan linguistik komputasi [7].

### 2.4 Preprocessing

*Preprocessing* merupakan proses menyiapkan data mentah untuk dianalisis atau digunakan dalam model pembelajaran mesin. Ini adalah langkah penting dalam proses pembelajaran mesin, karena kualitas dan karakteristik data dapat memengaruhi performa model secara signifikan. Proses tahap *preprocessing* dilakukan guna pengolahan data mentah menjadi *set* data yang sudah jadi untuk memilih data yang akan diproses dalam dokumen [8].

Dalam Gambar 2, akan melakukan serangkaian proses untuk menyaring, menghapus, dan memperbaiki kata-kata. Tujuan dilakukan *Preprocessing* ini yaitu menghasilkan teks yang lebih tertata dan terstruktur dari data *tweet*, yang dikenal sebagai *clean text*. Proses yang terjadi dalam tahap *Preprocessing* ini meliputi: mengubah huruf kecil (*Case Folding*), pembersihan (*Cleansing*), perbaikan kata (*normalization*), penghapusan kata penghubung (*Stop Word*), *Stemming*, dan *Filtering*.



**Gambar 2.** Preprocessing

Berikut penjelasan tahapan *Preprocessing*-nya:

- a. *Cleansing*  
 Proses *cleansing* dilakukan dengan tujuan agar data bersih dari elemen yang dapat merusak model serta mempercepat proses komputasi. Proses ini dilakukan dengan menghilangkan kata-kata diantaranya yakni url, *username* (@), *hashtag* (#), *emoticon*, kata yang dianggap spam, filter kata alay dan *case folding* [9].
- b. *Case Folding*  
*Case Folding* adalah mengubah semua huruf yang terdapat pada dokumen menjadi huruf kecil, hanya huruf yang terdapat dalam alphabet a sampai z saja yang diterima dan karakter lainnya yang dihilangkan [6].
- c. *Normalization*  
 Proses ini digunakan untuk perbaikan kata-kata yang masih salah dalam pengejaan atau merupakan singkatan berbentuk tertentu [10].
- d. Menghapus *Stopword*  
*Stopword* adalah kata yang sering muncul dalam bahasa yang tanpa memiliki banyak informasi untuk aplikasi *text mining*. Misalnya, kata-kata “a,” “an,” dan “the” adalah kata-kata yang umum muncul yang memberikan sedikit informasi tentang isi dokumen yang sebenarnya [6].
- e. *Stemming*  
 Dalam banyak kasus, *stemming* mengacu pada ekstraksi *root* umum dari kata-kata, dan *root* yang diekstraksi bahkan mungkin bukan kata itu sendiri. Misalnya, *root* umum dari *hoping* dan *hope* adalah *hop*. Kekurangannya adalah kata *hop* memiliki arti dan penggunaan yang berbeda. Oleh karena itu, meskipun *stemming* biasanya meningkatkan *recall* dalam pencarian dokumen, tetapi terkadang dapat sedikit memperburuk *precision* [6].
- f. *Filtering*  
 Dalam tahapan *Filtering* ini dilakukan untuk melakukan penyaringan kalimat dalam *tweet* atau penghapusan kata [11].

## 2.5 TF-IDF

Metode *Term Frequency-Inverse Document Frequency* adalah cara untuk memberi bobot pada hubungan sebuah kata (*term*) dengan sebuah dokumen. Metode ini menggabungkan dua konsep untuk menghitung bobot, yaitu frekuensi kemunculan suatu kata dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut [12]. Secara matematis, bobot TF-IDF dapat dihitung menggunakan Persamaan 1.

$$TF(t) = \sum F_{f,d} \quad (1)$$

Keterangan:

$F$  = frekuensi kemunculan

$t$  = kata

$d$  = dokumen

Dilanjutkan dengan mencari nilai IDF dengan Persamaan 2.

$$IDF(t) = \log \frac{D}{DF(t)} \quad (2)$$

Keterangan:

$D$  = total jumlah dokumen

$DF(t)$  = total dokumen yang mengandung frekuensi kemunculan kata  $t$

Setelah kedua nilai didapatkan, dilakukan Persamaan nilai TF-IDF dengan Persamaan 3.

$$TF - IDF(t) = TF(t) \times IDF(t) \quad (3)$$

Demikianlah beberapa Persamaan yang digunakan untuk menghitung dengan menggunakan metode TF-IDF [11].

## 2.6 Naive Bayes

Pada metode *Naive Bayes Classifier* menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses terhadap sampel data yang sedapat mungkin dapat menjadi representasi data tersebut. Pada probabilitas prior bagi tiap kategori berdasarkan sampel data [10].

$$P(H|X) = \frac{P(X|H) * P(H)}{P(X)} \quad (4)$$

Keterangan:

- $X$  = Data dengan class yang belum diketahui
- $H$  = Hipotesis data menggunakan suatu class yang spesifik
- $P(H|X)$  = Probabilitas hipotesis  $H$  berdasarkan kondisi  $X$  (Parteori Probabilitas)
- $P(H)$  = Probabilitas hipotesis  $H$  (Prior Probabilitas)
- $P(X|H)$  = Probabilitas  $X$  berdasarkan kondisi pada hipotesis  $H$
- $P(X)$  = Probabilitas  $X$

Demikian dari Persamaan 4 yang digunakan untuk menghitung dengan menggunakan metode *Naive Bayes* [12].

## 2.7 Confusion Matrix

Pengukuran kinerja algoritma klasifikasi pada penelitian ini yaitu dengan menggunakan *confusion matrix*. *Confusion matrix* menunjukkan hasil identifikasi antara jumlah data prediksi yang benar dari jumlah data prediksi yang salah dibandingkan dengan fakta yang dihasilkan [12]. Tabel *confusion matrix* dapat dilihat pada Tabel 1.

**Tabel 1.** *Confusion Matrix*

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	TP	FP	PNe
	Negatif	FN	TN	NNe
	Netral	NeP	NeN	TNe

Keterangan:

- TP (True Positif) = Jumlah dokumen dari kelas Positif yang benar diklasifikasikan sebagai kelas Positif.
- TN (True Negatif) = Jumlah dokumen dari kelas Negatif yang benar diklasifikasikan sebagai kelas Negatif.
- FP (False Positif) = Jumlah dokumen dari kelas Negatif yang salah diklasifikasikan sebagai kelas Positif.
- FN (False Negatif) = Jumlah dokumen dari kelas Positif yang salah diklasifikasikan sebagai kelas Negatif.
- PNe (Positif Netral) = Jumlah dokumen dari kelas Positif yang diklasifikasikan sebagai kelas Netral.
- NNe (Negatif Netral) = Jumlah dokumen dari kelas Negatif yang diklasifikasikan sebagai kelas Netral.
- NeP (Netral Positif) = Jumlah dokumen dari kelas Netral yang diklasifikasikan sebagai kelas Positif.
- NeN (Netral Negatif) = Jumlah dokumen dari kelas Netral yang diklasifikasikan sebagai kelas Negatif.
- TNe (True Netral) = Jumlah dokumen dari kelas Netral yang benar diklasifikasikan sebagai kelas Netral.

Parameter yang digunakan untuk performansi klasifikasi yaitu *accuracy*, *precision*, dan *recall*. Akurasi merupakan rasio kinerja observasi yang diprediksi dengan benar dari total observasi. Akurasi dapat dihitung menggunakan Persamaan 5.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

*Precision* merupakan rasio observasi positif yang diprediksi dengan benar dari total observasi positif yang diprediksi. Untuk mencari nilai *precision* digunakan Persamaan 6.

$$precision = \frac{TP}{TP + FP} \quad (6)$$

Sedangkan *recall*, bias disebut *sensitivity* adalah rasio pengamatan positif yang diprediksi dengan benar untuk semua pengamatan di kelas yang sebenarnya. Nilai *recall* dapat dihitung menggunakan Persamaan 7.

$$recall = \frac{TP}{TP + FN} \quad (7)$$

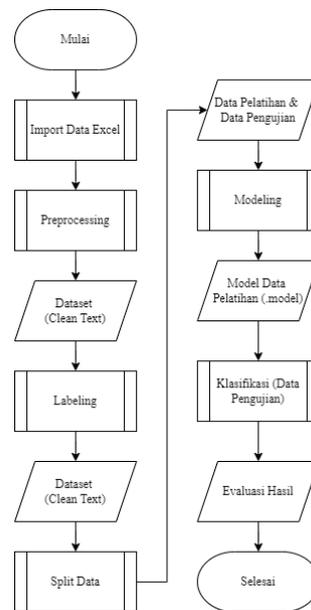
Demikianlah beberapa Persamaan yang digunakan untuk menghitung dengan menggunakan metode *Confusion matrix* [12].

### 3. HASIL PEMBAHASAN

#### 3.1 Penerapan Metode

Untuk penelitian ini, sebagai pembangunan sentimen analisis melibatkan serangkaian tahapan. Setiap tahapan tersebut mencerminkan proses dan desain yang diterapkan dalam penelitian, mulai dari awal hingga aplikasi berjalan dengan baik.

Dalam Gambar 3, menjelaskan mengenai *flowchart* terhadap keseluruhan sistem pada program sentimen analisis ini, pertama siapkan terlebih dahulu dataset yang berisi data hasil *crawling* menggunakan *twitter-harvest*, selanjutnya dataset akan dimasukkan melalui menu *Import Data Excel*, data tersebut akan masuk ke dalam baris data (*database*) di *MySQL*. Kemudian, data dalam *database* akan melalui tahap *Preprocessing*, *dataset* teks bersih tersebut akan diberi label untuk menentukan apakah labelnya positif, negatif, atau netral pada tahapan *Labeling*. Data yang sudah diberi label ini akan dibagi menjadi 80% untuk data pelatihan dan 20% untuk data pengujian menuju *Split Data*. Setelah itu, data ini akan di proses dalam tahap *Modeling* untuk menghasilkan *Confusion matrix* dan Klasifikasi menggunakan metode *Naive Bayes*.



Gambar 3. Proses Keseluruhan Sistem

Berikut merupakan tahapan dari proses keseluruhan sistem pada subjudul berikut ini:

#### 3.2 Import Data Excel

Proses *import data* dari file *Excel* yang dimulai dari persiapan data dalam format *xlsx*, *xls*, atau *csv*. Data tersebut di *import* dengan menggunakan fitur *import data Excel* dari program *website*-nya, seterusnya data yang telah di *import* akan di simpan ke dalam *database MySQL*.

#### 3.3 Preprocessing

Setiap proses dalam tahap *Preprocessing* dijelaskan secara detail di sini, mulai dari pembersihan data, pemisahan teks (*tokenization*), normalisasi teks, hingga penghapusan kata-kata tidak penting (*stopword removal*) dan *stemming*.

##### a. Cleansing

Dalam proses *cleansing* melibatkan pembersihan dan penghapusan teks serta simbol atau tanda baca yang tidak diperlukan. Tahapan ini mencakup beberapa langkah-langkah penting, yaitu: menghapus angka, menghapus mention, menghapus emoji, menghapus tautan, dan menghapus karakter khusus. Berikut merupakan contohnya pada Tabel 2.

Tabel 2. Proses *Cleansing*

Tweet	Cleansing
@abulmuzaffar10 Dissenting opinion hakim MK juga bukan contoh yang bagus buat nunjukin the power of anak abah itu pressurenya ya dari elit juga besarnya. Amicus curiae Megawati jadi game changer sesungguhnya	Dissenting opinion hakim MK juga bukan contoh yang bagus buat nunjukin the power of anak abah itu pressurenya ya dari elit juga besarnya Amicus curiae Megawati jadi game changer sesungguhnya

b. *Case Folding*

Proses *Case Folding* merupakan proses untuk mengubah semua teks menjadi huruf kecil untuk memudahkan dalam melakukan proses tokenisasi. Misalnya, kata “Amicus Curiae” akan diubah menjadi “amicus curiae”, sehingga menghilangkan perbedaan yang disebabkan oleh penggunaan huruf kapital. Berikut merupakan contohnya pada Tabel 3.

**Tabel 3.** Proses *Case Folding*

<i>Cleansing</i>	<i>Case Folding</i>
Dissenting opinion hakim MK juga bukan contoh yang bagus buat nunjukin the power of anak abah itu pressurenya ya dari elit juga besarnya Amicus curiae Megawati jadi game changer sesungguhnya	dissenting opinion hakim mk juga bukan contoh yang bagus buat nunjukin the power of anak abah itu pressurenya ya dari elit juga besarnya amicus curiae megawati jadi game changer sesungguhnya

c. *Tokenisasi*

Pada proses ini, setiap kalimat dipecah menjadi kata-kata individu. Tujuan dari pemisahan kata ini untuk memungkinkan dalam pengolahan setiap kata secara terpisah pada tahap berikutnya, yang memungkinkan analisis menjadi lebih mendalam dan efektif. Berikut ini merupakan contohnya pada Tabel 4.

**Tabel 4.** Proses *Tokenisasi*

<i>Case Folding</i>	<i>Tokenisasi</i>
dissenting opinion hakim mk juga bukan contoh yang bagus buat nunjukin the power of anak abah itu pressurenya ya dari elit juga besarnya amicus curiae megawati jadi game changer sesungguhnya	“dissenting” “opinion” “hakim” “mk” “juga” “bukan” “contoh” “yang” “bagus” “buat” “nunjukin” “the” “power” “of” “anak” “abah” “itu” “pressurenya” “ya” “dari” “elit” “juga” “besarnya” “amicus” “curiae” “megawati” “jadi” “game” “changer” “sesungguhnya”

d. *Normalisasi*

Di proses normalisasi, setelah kalimat yang dipecah menjadi kata-kata individu setelah melakukan tokenisasi akan diperiksa dengan cermat untuk mendeteksi adanya singkatan atau kata-kata informal. Berikut ini merupakan contohnya pada Tabel 5.

**Tabel 5.** Proses *Normalisasi*

<i>Tokenisasi</i>	<i>Normalisasi</i>
“dissenting” “opinion” “hakim” “mk” “juga” “bukan” “contoh” “yang” “bagus” “buat” “nunjukin” “the” “power” “of” “anak” “abah” “itu” “pressurenya” “ya” “dari” “elit” “juga” “besarnya” “amicus” “curiae” “megawati” “jadi” “game” “changer” “sesungguhnya”	“dissenting” “opinion” “hakim” “mahkamah konstitusi” “juga” “bukan” “contoh” “yang” “bagus” “buat” “nunjukin” “the” “power” “of” “anak” “abah” “itu” “pressurenya” “ya” “dari” “elit” “juga” “besarnya” “amicus” “curiae” “megawati” “jadi” “game” “changer” “sesungguhnya”

e. *Stopword Removal*

Proses penghapusan *stopword*, kata-kata yang kurang penting dan sering muncul dalam kalimat akan dihilangkan. Kata-kata yang akan dihapus ditandai dengan huruf tebal. Proses penghapusan tersebut menggunakan *Pustaka sastrawi*. Contoh dari kata tersebut adalah ‘dan’, ‘yang’, ‘di’, dan sebagainya. Berikut ini merupakan contohnya pada Tabel 6.

**Tabel 6.** Proses *Stopword Removal*

<i>Normalisasi</i>	<i>Stopword Removal</i>
“dissenting” “opinion” “hakim” “mahkamah konstitusi” “juga” “bukan” “contoh” “yang” “bagus” “buat” “nunjukin” “the” “power” “of” “anak” “abah” “itu” “pressurenya” “ya” “dari” “elit” “juga” “besarnya” “amicus” “curiae” “megawati” “jadi” “game” “changer” “sesungguhnya”	“dissenting” “opinion” “hakim” “mahkamah konstitusi” “contoh” “bagus” “nunjukin” “the” “power” “anak” “abah” “pressurenya” “elit” “besar” “amicus” “curiae” “megawati” “game” “changer” “sungguh”

f. *Stemming*

Tahap berikutnya mengubah kata-kata dengan imbuhan menjadi bentuk dasarnya. Sebagai contoh, kata ‘opinion’ akan menjadi ‘opini’, kata ‘pendukung’ akan menjadi ‘dukung’, dan kata ‘menolak’ akan menjadi ‘tolak’. Untuk menjamin akurasi dan konsistensi pengubahan kata, proses *stemming* ini dilakukan dengan menggunakan *Pustaka Sastrawi*. Berikut ini merupakan contohnya pada Tabel 7.

**Tabel 7.** Proses *Stemming*

<i>Stopword Removal</i>	<i>Stemming</i>
“dissenting” “opinion” “hakim” “mahkamah konstitusi” “contoh” “bagus” “nunjukin” “the” “power” “anak” “abah” “pressurennya” “elit” “besar” “amicus” “curiae” “megawati” “game” “changer” “sungguh”	dissenting opinion hakim mahkamah konstitusi contoh bagus nunjukin the power anak abah pressurennya elit besar amicus curiae megawati game changer sungguh

g. *Filtering*

Proses *filtering*, tujuannya untuk menghapus kata-kata dan simbol yang tidak relevan. Ini memastikan bahwa data yang digunakan untuk analisis adalah data yang bersih, relevan, dan tidak ada duplikasi, yang dapat memberikan hasil pemodelan sentimen yang lebih akurat dan signifikan.

**3.4 Labeling**

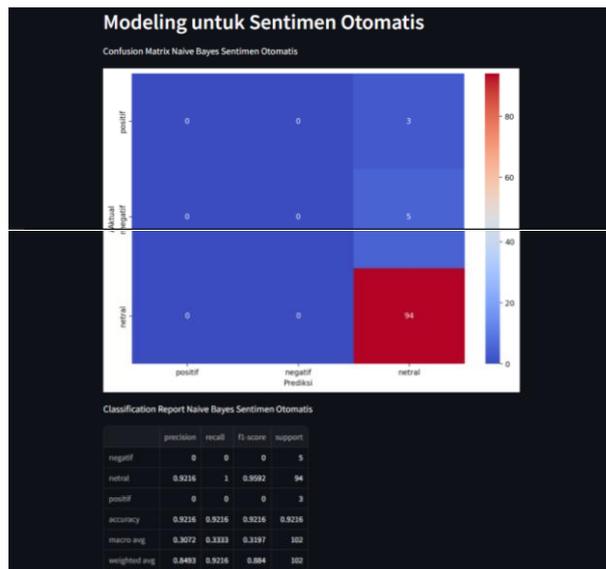
Dalam tahapan ini, teks dari hasil *preprocessing* dianalisis menggunakan VADER *Lexicon* dan evaluasi oleh ahli pakar untuk menentukan kelas sentimen positif, negatif, dan netral. Dalam penelitian ini, 509 *tweet* dianalisis menggunakan VADER, menghasilkan 482 *tweet* untuk kelas netral, 10 *tweet* untuk kelas positif, dan 17 *tweet* untuk kelas negatif. Selain itu, penilaian sentimen oleh ahli pakar yang menyarankan untuk menggunakan aplikasi *Nvivo* yang berguna untuk memastikan peningkatan akurasi yang signifikan. Dalam penelitian ini, memberikan label pada 509 *tweet*, dengan hasil 500 *tweet* untuk kelas netral, 3 *tweet* untuk kelas positif, dan 6 *tweet* untuk kelas negatif. Pendekatan ini yang menggabungkan analisis otomatis dan evaluasi ahli pakar, bertujuan untuk mencapai hasil yang lebih akurat dan dapat diandalkan dalam sentimen analisis.

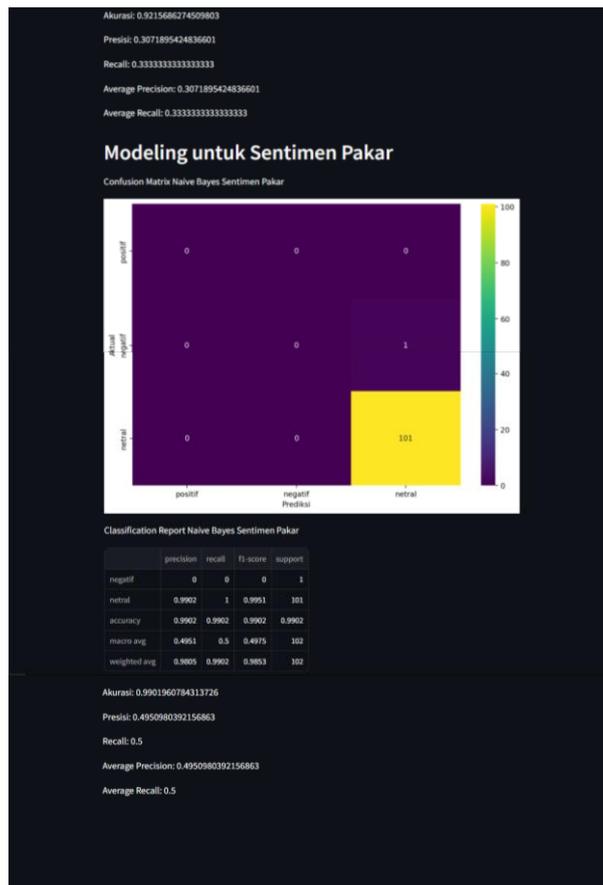
**3.5 Split Data**

Dalam tahapan *split data* ini, dibagi menggunakan metode stratified sampling. Berdasarkan persentase label dari setiap *tweet*, dihitung persentase untuk label positif, negatif, dan netral. Dengan 80% data digunakan untuk pelatihan (*training*) dan 20% digunakan untuk data pengujian (*testing*). Dalam penelitian ini, teknik *stratified sampling* digunakan untuk membagi 509 *tweet* dengan rasio 80:20. Hasilnya, 407 *tweet* digunakan untuk data pelatihan dan 102 *tweet* sebagai data pengujian. Kedua *set* data tersebut kemudian di simpan dalam basis data yang kemudian di proses ke tahap pemodelan data.

**3.6 Modeling**

Dalam tahapan ini, menggunakan metode TF-IDF untuk mengubah teks menjadi fitur numerik dan metode *Naive Bayes* sebagai pemberian hasil sentimen untuk kelas sentimen otomatis dan kelas sentimen pakar dari data yang sudah melalui proses *split data*. Untuk hasil dari proses *Modeling* ini, akan menampilkan *confusion matrix*, laporan klasifikasi, nilai akurasi, nilai presisi, nilai *recall* nilai rata-rata presisi, dan nilai rata-rata *recall*. Tampilan ini dapat dilihat pada Gambar 4.





**Gambar 4.** Proses *Modeling*

### 3.7 Klasifikasi

Pengujian klasifikasi merupakan tahapan penting dalam melakukan pengembangan sistem, bertujuan untuk mengevaluasi, menganalisis, dan memahami berapa tingkat akurasi atau kesesuaian hasil yang dicapai oleh sistem yang telah dirancang. Jumlah data sentimen yang dikumpulkan sebanyak 509 *tweet* dianalisis yang menggunakan metode VADER, menghasilkan 482 *tweet* kelas netral, 10 *tweet* kelas positif, dan 17 *tweet* kelas negatif. Selain itu, penilaian sentimen oleh ahli pakar yang menyarankan untuk menggunakan aplikasi *Nvivo*, memberikan label pada 509 *tweet*, dengan hasil 500 *tweet* kelas netral, 3 *tweet* kelas positif, dan 6 *tweet* kelas negatif. Dalam pengujian ini, tidak hanya mengukur seberapa baik sistem yang memenuhi spesifikasi yang diharapkan tetapi juga mengidentifikasi area yang memerlukan perbaikan. Untuk menunjukkan persentase dari masing-masing kelas sentimen positif, negatif, dan netral pada data sentimen otomatis dan data sentimen pakar, dapat menggunakan perhitungan dari Persamaan 8.

$$S = \frac{S_i}{S_{total}} \times 100\% \quad (8)$$

Keterangan:

- $S$  = Persentase dari kelas sentimen tertentu (positif, negatif, dan netral).
- $S_i$  = Jumlah *tweet* dari kelas sentimen tertentu (positif, negatif, atau netral).
- $S_{total}$  = Jumlah total data *tweet*.

Dalam pengujian yang dilakukan, algoritma *Naive Bayes* menunjukkan bahwa data sentimen otomatis terdapat sentimen positif dalam data itu mencapai 1,96%, sentimen negatif sebesar 3,34%, dan sentimen netral sebesar 94,69%. Sedangkan, pada data sentimen pakar terdapat sentimen positif dalam data itu mencapai 0,59%, sentimen negatif sebesar 1,18%, dan sentimen netral sebesar 98,23%.

Dari hasil ini, dapat disimpulkan bahwa *Naive Bayes* memiliki keterbatasan signifikan dalam memprediksi kelas positif dan kelas negatif. Meskipun menunjukkan performa baik dalam memprediksi kelas netral. Perlu adanya peningkatan atau penggunaan algoritma lain untuk mencapai hasil yang lebih dalam prediksi seluruh kelas.

#### 4. KESIMPULAN

Setelah melakukan pengujian dan evaluasi terhadap aplikasi yang dikembangkan menggunakan dataset dan Algoritma yang diajukan, dapat ditarik kesimpulan bahwa:

- a. Penelitian ini berhasil menciptakan *prototype* untuk klasifikasi sentimen terhadap pengajuan *Amicus Curiae* ke Mahkamah Konstitusi berbasis *web*.
- b. Berdasarkan analisis terhadap 509 tweet pada rentang waktu 18 April sampai 27 Mei 2024, pada data sentimen otomatis terdapat sentimen positif dalam data itu mencapai 1,96%, sentimen negatif sebesar 3,34%, dan sentimen netral sebesar 94,69%. Sedangkan, pada data sentimen pakar terdapat sentimen positif dalam data itu mencapai 0,59%, sentimen negatif sebesar 1,18%, dan sentimen netral sebesar 98,23%. Dapat diketahui bahwa mayoritas masyarakat Indonesia memiliki pandangan netral terhadap pengajuan *Amicus Curiae* ke Mahkamah Konstitusi pada kasus umum di Indonesia tahun 2024.
- c. Dalam penggunaan pembobotan kata Algoritma *Naive Bayes* dalam melakukan sentimen analisis telah memberikan hasil nilai pengujian dan evaluasi untuk data sentimen otomatis yaitu akurasi sebesar 92,2%, presisi sebesar 0,92, dan *recall* sebesar 1,0. Sedangkan, untuk data sentiment pakar yaitu akurasi sebesar 0,99, presisi sebesar 0,99, dan *recall* sebesar 1,0.

#### DAFTAR PUSTAKA

- [1] T. Gunawan, A. A. H. Putra, A. Faturrohman, and N. A. Rakhmawati, "Analisis Sentimen Pembacaan Putusan MK Terkait Pilpres 2024 pada Aplikasi X dengan Metode Naive Bayes Classifier," *ResearchGate*, vol. 1, no. 1, pp. 0–8, 2024, [Online]. Available: <https://www.researchgate.net/publication/381830593>
- [2] W. Gata and A. Bayhaqy, "Analysis Sentiment about Islamophobia when Christchurch Attack on Social Media," *Telkomika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 4, pp. 1819–1827, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14179.
- [3] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [4] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [5] R. P. I. Putra, M. Akbar, and R. Amalia, "Analisis Sentimen Masyarakat Terhadap Kinerja Persatuan Sepakbola Seluruh Indonesia Menggunakan Metode Backpropagation," *J. Inf. Technol. Ampera*, vol. 1, no. 2, pp. 106–118, 2020, doi: 10.51519/journalita.volume1.issue2.year2020.page106-118.
- [6] I. Sari and D. Indarti, "Text Mining: Praktik Klasifikasi dan Pemodelan Topik dengan Python," in *Uwais Inspirasi Indonesia*, Pertama., Ds. Sidoarjo, Kec. Pulung, Kab. Ponorogo: Uwais Inspirasi Indonesia, 2022, p. vi + 112 halaman.
- [7] Syahriani, A. A. Yana, and T. Santoso, "Sentiment analysis of facebook comments on indonesian presidential candidates using the naive bayes method," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, pp. 1–11, 2020, doi: 10.1088/1742-6596/1641/1/012012.
- [8] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN," *J. KomtekInfo*, vol. 10, no. 1, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [9] S. Mujahidin, B. Prasetio, and M. C. C. Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naive bayes," *Voteteknika (Vocational Tek. Elektron. dan Inform.*, vol. 10, no. 3, pp. 17–24, 2022, doi: 10.24036/voteteknika.v10i3.118299.
- [10] R. Noviana and I. Rasal, "Penerapan Algoritma Naive Bayes Dan SVM Untuk Analisis Sentimen Boy Band BTS Pada Media Sosial Twitter," *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
- [11] D. Oktavia, Y. R. Ramadhan, and M. Minarto, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [12] M. Lestandy, A. Abdurrahim, and L. Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naive Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.