

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Berita merupakan suatu bentuk laporan pendapat seseorang atau kelompok maupun peristiwa tentang temuan baru disegala aspek yang dianggap penting bagi wartawan yang bertujuan untuk ditampilkan di media (Anton Mabruri KN, 2018). Berita dulu hanya muncul di surat kabar, radio, dan televisi. Namun, dengan kemajuan teknologi internet dan peningkatan jumlah pengguna internet di Indonesia, menurut data Kementerian Komunikasi dan Informatika Republik Indonesia, pada tahun 2023, mencapai 212,9 juta orang, setara dengan 77% populasi Indonesia pada Januari 2023. Akibatnya, sumber informasi berita kini lebih dominan berasal dari media *online*. (Kominfo, 2023). Media *online* seperti berita online, memiliki unsur-unsur seperti *headline*, *lead*, tubuh berita, dan foto (Azhar&Sari, 2018). Sebagai contoh, Kompas.tv, sebagai salah satu platform berita online, memiliki struktur judul berita yang mampu menjelaskan dengan efektif konten yang akan dibahas dalam berita tersebut. Dalam proses pengklasifikasian jenis berita, judul berita dianggap sebagai kriteria utama yang dapat memandu pembaca untuk menentukan relevansi dan pentingnya suatu topik berita.

Klasifikasi merupakan teknik dalam pembelajaran mesin yang bertujuan memprediksi kelas atau label suatu data dengan memanfaatkan fitur-fiturnya. Proses klasifikasi pada dasarnya memisahkan data menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membentuk model klasifikasi, sementara data uji digunakan untuk menguji performa model tersebut dalam memprediksi kelas atau label dari data yang belum pernah dilihat sebelumnya (Provost & Fawcett, 2019). Proses klasifikasi melibatkan pemeriksaan dan analisis dokumen teks pra-klasifikasi untuk membentuk model yang dapat digunakan dalam mengklasifikasikan dokumen teks baru yang belum memiliki kategori. Pengklasifikasian teks dapat dilakukan melalui bidang data mining yang dikenal sebagai *text mining*. *Text mining* merupakan suatu teknik dalam ilmu data yang bertujuan untuk mengekstrak informasi berharga dari teks. Dalam konteks analisis

sentimen, text mining dapat mengidentifikasi sentimen dari suatu pernyataan. Teknik ini digunakan untuk mengidentifikasi pola dan hubungan yang terdapat dalam data teks (Silge, J., & Robinson, 2019). Identifikasi yang dimaksud salah satunya adalah klasifikasi sentimen.

Analisis sentimen merupakan penelitian yang tergolong dalam *text mining* yang memiliki fungsi mengklasifikasikan data atau dokumen dalam bentuk tekstual atau teks sebagai kalimat opini berbasis sentimen (Rizal Setya Perdana, 2018). Menurut penelitian yang dilakukan oleh Delfariyadi et al. (2022), sentimen dapat diklasifikasikan menggunakan metode *Naïve Bayes Classifier*. (Delfariyadi et al., 2022). Metode *Naïve Bayes Classifier* telah terbukti efektif dalam pengklasifikasian sentimen dengan kecepatan pemrosesan dan pengklasifikasian data yang tinggi (Miftakhurahmat et al., 2023). Metode ini menunjukkan performa yang baik dalam mengidentifikasi sentimen dari teks atau data yang diberikan. Namun, perlu diingat bahwa *Naïve Bayes Classifier* memiliki kekurangan. Salah satunya adalah ketidakberlakuannya ketika terdapat nilai probabilitas yang nol. Jika terjadi nilai nol, prediksi yang dihasilkan juga akan menjadi nol. Oleh karena itu, perlu dilakukan penanganan khusus saat menghadapi kasus-kasus di mana probabilitas nol mungkin terjadi untuk mencapai hasil yang lebih akurat dalam pengklasifikasian sentimen (Singh et al., 2022)

Dalam pengklasifikasian *text mining*, algoritma *Naïve Bayes Classifier* digunakan dalam penelitian yang berjudul “Klasifikasi Sentimen Judul Berita Pemberitaan COVID-19 Tahun 2021 pada Media DetikHealth” dengan total 399 judul berita sepanjang tahun 2021 ditemukan hasil judul berita dengan label netral sebanyak 147 data, label positif sebanyak 114 data, dan label negatif sebanyak 138 data. Dalam penelitian ini didapatkan juga bahwa implementasi dari algoritma *Naïve Bayes Classifier*, persentase akurasi yang diperoleh sebesar 72.5% (Delfariyadi et al., 2022).

Dari latar belakang diatas, maka penelitian ini menjadi sebuah Tugas Akhir dengan judul “Penerapan Metode *Naive Bayes Classifier* Untuk Klasifikasi

Sentimen Judul Berita Kompas.Tv Kota Semarang”, dengan menggunakan dua kelas yaitu positif dan negatif, dengan adanya dua kelas maka dapat dianalisa hasil dari klasifikasinya dan diharapkan penelitian ini mampu meningkatkan performa akurasi pada penelitian sebelumnya dengan metode *Naïve Bayes Classifier*.

### 1.2. Rumusan Masalah

Dengan merujuk pada konteks yang telah diuraikan, rumusan masalah yang diajukan adalah sejauh mana penerapan serta kinerja metode Klasifikasi Naive Bayes terhadap pengelompokan sentimen pada judul berita online di Kompas TV, khususnya di wilayah Kota Semarang.

### 1.3. Batasan Masalah

Batasan masalah pada penelitian dengan judul "Penerapan Metode *Naive Bayes Classifier* Untuk Klasifikasi Sentimen Judul Berita Kompas.Tv Kota Semarang" adalah sebagai berikut:

1. Penelitian ini hanya fokus pada *headline* berita Kompas.tv Kota Semarang pada kurun waktu tahun 2023.
2. Penelitian ini hanya mengklasifikasikan sentimen dalam *headline* berita Kompas.tv Kota Semarang menjadi dua kategori: positif, dan negatif.
3. Metode Naive Bayes Classifier yang diterapkan dalam penelitian ini hanya memanfaatkan fitur-fitur yang relevan untuk mengklasifikasikan sentimen dalam judul berita.
4. Penelitian ini tidak membahas mengenai pengaruh konteks dan penggunaan bahasa dalam *headline* berita terhadap klasifikasi sentimen.

### 1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk mengevaluasi kemampuan metode Naive Bayes Classifier dalam mengidentifikasi dan mengklasifikasikan sentimen positif dan negatif yang terdapat dalam judul-judul berita. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan mengenai efektivitas metode Naive Bayes

Classifier dalam menganalisis sentimen pada judul berita Kompas.TV di wilayah Kota Semarang.

### **1.5. Manfaat Penelitian**

Manfaat dari penelitian dengan judul "Penerapan Metode *Naive Bayes Classifier* Untuk Klasifikasi Sentimen Judul Berita Kompas.Tv Kota Semarang" adalah sebagai berikut:

1. Bagi akademik:
  - a. Memperluas pengetahuan dan pemahaman terkait penggunaan algoritma *Naive Bayes Classifier* dalam klasifikasi teks pada judul berita online Kompas.TV di Kota Semarang..
  - b. Sebagai referensi bagi peneliti atau akademisi yang ingin melakukan penelitian serupa pada media berita lainnya.
  - c. Memberikan kontribusi dalam pengembangan metode klasifikasi sentimen pada *headline* berita dengan metode *Naive Bayes Classifier*.
2. Bagi masyarakat:
  - a. Mengetahui sentimen yang terkandung dalam judul berita online Kompas.TV Kota Semarang.
  - b. Membantu masyarakat dalam memahami dan menafsirkan judul berita dengan lebih baik, dengan memperhatikan sentimen yang terkandung di dalamnya.
  - c. Memungkinkan masyarakat untuk menjadi lebih kritis terhadap berita yang mereka konsumsi, karena mereka dapat melihat sentimen yang terkandung dalam judul berita tersebut.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Studi

Dalam studi mengenai klasifikasi sentimen terhadap media daring dalam konteks Pemilihan Presiden Indonesia tahun 2019, metode Naïve Bayes Classifier diimplementasikan dengan judul "Analisis Sentimen Pada Media Daring Tentang Pemilihan Presiden Indonesia Tahun 2019 Menggunakan Metode Naïve Bayes". Hasil penelitian menunjukkan tingkat akurasi sebesar 80% untuk portal berita Detik dan 70% untuk Kompas, dengan pembagian dataset menjadi 90% untuk data pelatihan dan 10% untuk data uji (Jaya & Muslim, 2019).

Hasil penelitian yang berjudul "Text Mining untuk Mengklasifikasi Judul Berita Online: Studi Kasus Radar Banjarmasin dengan Metode Naïve Bayes" dengan menggunakan teknik *scraping* html untuk mendapatkan data. Sehingga terkumpul 400 data yang dibagi menjadi 4 kelas, yaitu banua, bisnis, hukum, dan kriminal. Dalam penelitian ini, data akan melalui beberapa proses yakni, *text preprocessing*, pengklasifikasian, dan pengujian metode. Penelitian ini mendapatkan hasil bahwa penggunaan metode *Naïve Bayes Classifier* menghasilkan akurasi sebesar 78,75%, dengan recall sebesar 80,56% dan precision sebesar 78,75% (Sholih 'afif et al., 2021).

Dalam penelitian sebelumnya tentang klasifikasi sentimen pada judul berita mengenai Covid-19 di Indonesia dengan metode Naïve Bayes Classifier, fokus penelitian kali ini adalah data judul berita yang diambil dari media daring DetikHealth. Total data yang digunakan sebanyak 399 judul berita selama tahun 2021. Data tersebut dibagi menjadi tiga kategori sentimen, yaitu sentimen netral dengan 147 data, sentimen positif sebanyak 114 data, dan sentimen negatif sejumlah 138 data. Setelah melalui proses text preprocessing dan penerapan algoritma Naïve Bayes Classifier, hasil penelitian menunjukkan tingkat akurasi sebesar 72,5% (Delfariyadi et al., 2022).

Dari penelitian yang berjudul “Analisis Sentimen Review Aplikasi Berita Online Pada Google Play Menggunakan Metode Algoritma Naive Bayes Classifier Dan Support Vector Machines”, dengan menggunakan *tools scraper similar* pada Google Chrome mendapatkan total 5894 data, kemudian data tersebut diolah melalui beberapa proses yakni *preprocessing*, TF-IDF, dan klasifikasi menghasilkan data ulasan sentimen positif sebanyak 4655 dan 421 ulasan sentimen negatif serta menghasilkan nilai akurasi dari Support Vector Machine (SVM) sebesar 94.06% sedangkan Naive Bayes sebesar 91.58% (Kusnia, 2022).

Menurut penelitian yang berjudul “Klasifikasi Judul Berita Online menggunakan Metode Support Vector Machine (SVM) dengan Seleksi Fitur Chi-square”. Penelitian ini melibatkan 2400 data yang diambil dari situs kompas.com dengan jumlah enam kelas yaitu olahraga, kesehatan, kuliner, berita nasional, teknologi, dan keuangan dengan setiap kelas memiliki 400 data. Data ini diambil dalam rentang tanggal 22 Februari 2023 sampai 7 April 2023. Penelitian ini menggunakan beberapa tahapan *pre-processing text*, *term weighting* menggunakan TF-IDF, seleksi fitur dengan chi-square, dan klasifikasi dengan SVM. Penggunaan SVM ini dilakukan untuk menilai judul berita online masuk ke dalam kategori mana. Hasil pengujian menunjukkan performa terbaik pada akurasi 93,06%, presisi 92,11%, recall 93,06%, dan f1-score 93,04%. Pencapaian ini diperoleh dengan mengatur threshold chi-square sebesar 80%, dan parameter SVM seperti kernel polynomial derajat 2,  $C=1$ ,  $\lambda=1$ , konstanta  $\gamma=0,01$ ,  $\epsilon=10^{-8}$ , dan maksimal iterasi sebanyak 10. (Rama et al., 2023).

Tabel 2.1 Penelitian Terkait

No	Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Jaya, Chandra Muslim, Kemas, 2019	Analisis sentimen pada platform media daring terkait Pemilihan Presiden Indonesia tahun 2019.	<i>Naïve Bayes Classifier</i>	Penelitian ini mencapai tingkat akurasi sebesar 80% pada portal berita Detik dan 70% pada Kompas, dengan pembagian dataset menjadi 90% untuk data latih dan 10% untuk data uji.
2.	Sholih 'afif, Muhammad Muzakir, Muhammad Al, Moh Iqbal Al Awalaieen, Ghifari, 2021	Analisis klasifikasi pada judul berita online Radar Banjarmasin.	<i>Naïve Bayes Classifier</i>	Dalam penelitian ini, diperoleh 400 data berita yang diklasifikasikan ke dalam empat kelas, yaitu banua, bisnis, hukum, dan kriminal. Hasil penelitian menunjukkan tingkat akurasi sebesar 78,75%, dengan recall mencapai 80,56% dan presisi sebesar 78,75%.
3.	Delfariyadi, Fahri Helen, Afrida Yuliawati, Susi, 2022	Analisis klasifikasi sentimen pada judul berita mengenai pemberitaan COVID-19 pada tahun 2021 di platform media daring DetikHealth.	<i>Naïve Bayes Classifier</i>	Dari hasil pengumpulan data 399 judul berita, dikelompokkan ke dalam tiga kategori sentimen, yakni 147 data dengan sentimen netral, 114 data dengan sentimen positif, dan 138 data dengan sentimen negatif. Dengan menerapkan metode Naive Bayes Classifier (NBC), diperoleh tingkat akurasi sebesar 72,5%.
4.	Kusnia, Ulfa, 2022	Klasifikasi ulasan terhadap aplikasi media berita online.	<i>Support Vector Machine &amp; Naïve Bayes Classifier</i>	Diperoleh data 4655 ulasan positif dan 421 ulasan negatif. Metode Support Vector Machine (SVM) mendapatkan akurasi 94.06%, sedangkan Naive Bayes mencapai akurasi 91.58%.
5.	Rama, Putu Putra, Bena	Klasifikasi judul berita online untuk	<i>Support Vector Machine</i>	Penelitian ini menghasilkan tingkat akurasi sebesar 93,06%, dengan presisi mencapai 92,11%, recall

	Perdana, Rizal Setya, 2023	membuat kategori berita		sebanyak 93,06%, dan f1-score mencapai 93,04%. Data yang digunakan dalam penelitian ini terdiri dari 2400 sampel yang diambil dari situs kompas.com. Enam kelas yang berbeda teridentifikasi, yakni olahraga, kesehatan, kuliner, berita nasional, teknologi, dan keuangan, dengan masing-masing kelas memiliki 400 data sampel.
--	----------------------------------	----------------------------	--	--

## 2.2. Tinjauan Pustaka

Sebelum memasuki isi tinjauan pustaka, penting untuk memahami konsep-konsep dasar yang terkait dengan penelitian ini. Tinjauan pustaka ini membahas beberapa konsep yang relevan, seperti *text mining*, metode klasifikasi seperti *Naïve Bayes Classifier*, analisis sentimen, *crawling data*, *preprocessing*, dan seleksi fitur. Pemahaman tentang konsep-konsep ini memberikan landasan teoritis yang kuat untuk penelitian ini dan membantu dalam memahami metode yang digunakan dalam analisis sentimen pada penelitian kali ini. Selanjutnya, tinjauan pustaka menjelaskan secara lebih rinci tentang masing-masing konsep dan penelitian terkait yang telah dilakukan sebelumnya.

### 2.2.1. Text Mining

Salah satu cabang dari data mining yang khusus berfokus pada analisis dan pemrosesan teks. Tujuan utama dari *text mining* adalah untuk menemukan pola atau bentuk unik dari kumpulan data berjenis teks yang memiliki volume besar. Dalam text mining, data teks diolah dan dianalisis untuk menghasilkan informasi yang dapat dimanfaatkan dalam berbagai bidang (Roufia, 2018).

### 2.2.2. Supervised Learning

Klasifikasi merupakan salah satu teknik yang digunakan dalam teknologi *big data*. Dimana dalam klasifikasi memiliki dua model klasifikasi yang umum digunakan, yaitu supervised learning dan unsupervised learning. Dalam penelitian ini, metode klasifikasi yang



digunakan adalah supervised learning, yang bertujuan untuk mengelompokkan dataset berdasarkan label atau kelas yang berbeda serta melakukan prediksi pada data yang ada (Kurniawan et al., 2022).

### **2.2.3. Analisis Sentimen**

Analisis sentimen merupakan penelitian yang tergolong dalam *text mining* yang memiliki fungsi mengklasifikasikan data atau dokumen dalam bentuk tekstual atau teks sebagai kalimat opini berbasis sentimen (Rizal Setya Perdana, 2018). Pada penerapannya, analisis sentimen memiliki fungsi utama untuk mengukur tingkat kepuasan masyarakat terhadap berbagai entitas seperti produk, layanan jasa, organisasi, dan tokoh masyarakat. Tujuan dari analisis sentimen ini adalah untuk memberikan evaluasi yang dapat meningkatkan kualitas entitas tersebut. Dalam penelitian analisis sentimen ini, terdapat dua kategori yaitu positif, dan negatif.

### **2.2.4. Crawling Data**

Dalam penelitian ini, pengumpulan data judul berita dilakukan melalui teknik crawling. Proses *crawling* data dilakukan dengan melakukan *sniffing* atau memonitor situs terkait untuk mendapatkan respons berupa *body* HTML. Selanjutnya, dilakukan parsing data pada *body* HTML tersebut. Setelah parsing dilakukan, dilakukan pengambilan data yang relevan untuk keperluan penelitian ini..

### **2.2.5. Preprocessing**

Dalam analisis data, *preprocessing* merupakan tahap yang penting untuk mengolah data teks yang awalnya tidak terstruktur menjadi lebih terstruktur. Tujuannya adalah agar proses analisis data dapat dilakukan dengan lebih akurat dan efisien. Terdapat empat tahap pada *preprocessing* ini, antara lain (Ferryawan et al., 2020):

- a. *Case folding* merupakan salah satu tahapan dalam mengubah data teks dalam dokumen menjadi huruf kecil. Hal ini dilakukan untuk

memastikan konsistensi dalam analisis data dan mempermudah proses pengolahan lebih lanjut.

- b. *Tokenizing* merupakan proses pemisahan *string* berdasarkan setiap kata yang membentuk urutan teks. Pada tahap ini, *string* dibagi menjadi bagian-bagian yang terdiri dari kata-kata individual, memungkinkan untuk melakukan analisis lebih lanjut pada setiap kata tersebut..
- c. *Filtering* adalah seleksi kata-kata penting dari hasil tokenisasi sebelumnya. Pada tahap ini, dilakukan pengambilan kata-kata yang dianggap relevan dan penting untuk analisis lebih lanjut.
- d. *Stemming* merupakan proses pengelompokan kata-kata yang memiliki kata dasar dan arti yang sama, meskipun memiliki bentuk yang berbeda karena adanya variasi imbuhan.

#### 2.2.6. Term Frequency Inverse Document Frequency

*Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan metode pembobotan yang mengintegrasikan konsep *Term Frequency* (TF) dan *Document Frequency* (DF). *Term Frequency* mengukur seberapa sering suatu kata muncul dalam suatu dokumen, sementara *Document Frequency* mengukur berapa banyak dokumen yang mengandung kata tersebut. TF-IDF menghasilkan nilai bobot untuk suatu kata dalam suatu dokumen. Bobot ini digunakan untuk menentukan tingkat pentingnya suatu kata dalam konteks dokumen tersebut.. Semakin tinggi bobot TF-IDF, semakin penting kata tersebut dalam dokumen tersebut.

Berikut merupakan rumus dari perhitungan TF-IDF:

$$\text{IDF}(w) = \log\left(\frac{N}{\text{DF}(w)}\right) \quad (1)$$

$$\text{TF-IDF}(w, d) = \text{TF}(w, d) * \text{IDF}(w) \quad (2)$$

Keterangan:

TF-IDF : bobot kata pada semua dokumen

w : kata

$d$  : dokumen

$TF(w,d)$  : frekuensi jumlah kemunculan kata ( $w$ ) pada dokumen( $d$ )

$N$  : jumlah semua dokumen

$DF(w)$  : jumlah dokumen ( $d$ ) yang didalamnya terdapat kata ( $w$ ).

### 2.2.7. Naïve Bayes Classification

*Naive Bayes Classifier* merupakan metode klasifikasi yang berasal dari teorema Bayes. Metode ini menggunakan probabilitas dan statistik untuk melakukan klasifikasi. Naive Bayes Classifier memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Asfi & Fitrianiingsih, 2020).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

Keterangan:

$X$  : Data pada kelas yang belum diketahui

$H$  : Hipotesis data  $X$  adalah kelas yang spesifik

$P(X)$  : Probabilitas dari  $X$

$P(H)$  : Probabilitas dari  $H$

$P(H|X)$  : Probabilitas  $H$  berdasar kondisi hipotesis  $X$

$P(X|H)$  : Probabilitas  $X$  berdasar kondisi hipotesis  $H$

Probabilitas  $H$  berdasarkan kondisi hipotesis  $X$  dihasilkan dari probabilitas  $X$  berdasarkan hipotesis  $X$ , probabilitas  $H$ , serta probabilitas  $X$ . Namun dalam pemakaiannya, rumus atau persamaan ini akan berubah menjadi:

$$P(C) = \frac{P(W_i, C) + 1}{\text{Count}(C) + |V|} \quad (4)$$

Keterangan:

$P(C)$  : Probabilitas dari  $C$

$P(W_i, C)$  : Probabilitas dari fitur  $W_i$  muncul dalam kelas  $C$

Count( $C$ ): Jumlah dokumen yang termasuk dalam kelas  $C$

$|V|$  : Jumlah total fitur-fitur yang ada

### 2.2.8. Confusion Matrix

*Confusion matrix* adalah merupakan metode evaluasi yang digunakan untuk mengukur kinerja dari suatu metode klasifikasi. Dengan cara yang sederhana, confusion matrix menyajikan informasi perbandingan antara hasil klasifikasi yang diberikan oleh sistem dengan hasil klasifikasi yang sebenarnya. (Roufia, 2018). Berikut merupakan rumus dari perhitungan *confusion matrix*:

#### a. Accuracy

Akurasi adalah ukuran yang menunjukkan sejauh mana sistem atau program mampu menganalisis data dengan benar. Berikut ini adalah rumus untuk menghitung akurasi.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} * 100\% \quad (5)$$

Keterangan:

- a. TP (*True Positive*), adalah jumlah data positif yang diklasifikasikan dengan benar oleh sistem.
- b. TN (*True Negative*), adalah jumlah data negatif yang diklasifikasikan dengan benar oleh sistem.
- c. FN (*False Negative*), adalah jumlah data negatif yang diklasifikasikan dengan salah oleh sistem.
- d. FP (*False Positive*), adalah jumlah data positif yang diklasifikasikan dengan salah oleh sistem.

### **b. Precision**

*Precision* adalah ukuran yang menunjukkan sejauh mana sistem atau program mampu mengukur ketepatan dalam menghasilkan informasi yang diinginkan oleh pengguna. Berikut ini adalah rumus untuk menghitung precision:

$$\text{Precision} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (6)$$

Keterangan:

- a. TN (*True Negative*), merupakan total data negatif yang mampu diklasifikasikan dengan benar oleh sistem
- b. FN (*False Negative*), merupakan total data negatif yang mampu diklasifikasikan dengan salah oleh sistem.

### **c. Recall**

*Recall* adalah ukuran yang menunjukkan sejauh mana sistem atau metode dapat mengembalikan informasi yang benar dari data yang ada. Berikut ini adalah rumus untuk menghitung recall:

$$\text{Recall} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (7)$$

Keterangan:

- a. TN (*True Negative*), merupakan total data negatif yang mampu diklasifikasikan dengan benar oleh sistem
- b. FP (*False Positive*), merupakan total data positif yang mampu diklasifikasikan dengan salah oleh sistem.

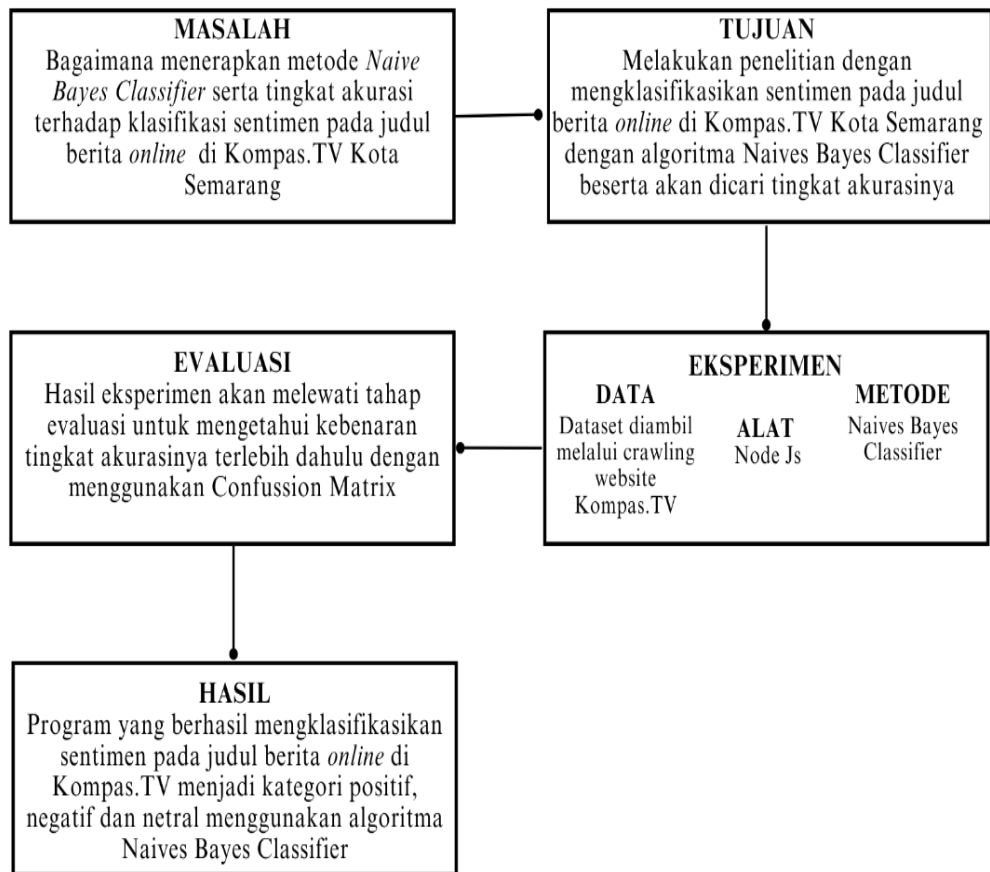
### **2.2.9. Node JS**

Node.js adalah *platform runtime* yang memungkinkan pengembangan aplikasi berbasis server-side menggunakan JavaScript. Dalam konteks *text mining* dan analisis sentimen, Node.js dapat digunakan

untuk memproses dan menganalisis teks secara efisien. Dalam penelitian ini, menggunakan beberapa *library* seperti Natural.

### **2.3. Kerangka Pemikiran**

Kerangka pemikiran yang digunakan dalam penelitian ini melibatkan serangkaian tahapan yang harus dijalani. Diantara langkah tersebut adalah latar belakang permasalahan, tujuan penelitian, tahap penelitian serta hasil penelitian. Dalam latar belakang penelitian, penulis melakukan riset terhadap klasifikasi sentimen berdasarkan judul berita pada media Kompas.TV Kota Semarang. Dalam penelitian ini, dilakukan klasifikasi sentimen berdasarkan judul berita dengan menggunakan metode Naive Bayes Classifier, serta pengukuran akurasi dari klasifikasi tersebut. Tahapan penelitian melibatkan penerapan metode Naive Bayes Classifier sebagai algoritma untuk mengklasifikasikan sentimen dalam judul berita media Kompas.TV di Kota Semarang. Pengolahan data dilakukan menggunakan bahasa pemrograman Node JS. Tujuan dari penelitian ini adalah menentukan tingkat akurasi dari metode yang digunakan. Pada tahap hasil penelitian, penulis menyimpulkan hasil dari riset yang dilakukan. Hasil ini memberikan gambaran tentang efektivitas metode yang digunakan dalam mengklasifikasikan sentimen berdasarkan judul berita. Selain itu, ditunjukkan juga tingkat akurasi dari metode yang digunakan. Untuk memudahkan pemahaman terhadap kerangka pemikiran dalam penelitian ini, berikut adalah diagram yang menjelaskan kerangka pemikirannya:



Gambar 2.1. Kerangka Pemikiran

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Sumber Data**

Data dalam penelitian ini bersumber dari judul berita online yang diambil dari situs web Kompas.TV Kota Semarang. Untuk mengumpulkan data tersebut, digunakan teknik *crawling* API dari situs web Kompas.TV. Proses pengumpulan data dilakukan melalui API yang telah disediakan oleh Kompas.TV. API ini memungkinkan akses terstruktur dan otomatis ke data judul berita yang terdapat di situs web tersebut. Dengan menggunakan teknik *crawling* API, peneliti dapat mengambil judul berita secara otomatis tanpa perlu melakukan akses manual ke setiap halaman berita. Oleh karena itu, dalam penelitian ini, data judul berita dari situs web Kompas.TV dimanfaatkan sebagai sumber data penelitian.

#### **3.2. Instrumen Penelitian**

Dalam penelitian ini, terdapat beberapa komponen yang diperlukan untuk mendukung kelancaran penelitian. Beberapa komponen tersebut meliputi:

##### **3.2.1. Kebutuhan Perangkat Lunak**

Beberapa perangkat lunak yang diperlukan meliputi:

- a. Sistem operasi yang digunakan pada perangkat penulis adalah Windows 11.
- b. Dalam menuliskan program atau sistem yang dibangun penulis menggunakan Visual Studio Code dan Node JS.
- c. Penulis menggunakan Google Chrome sebagai browser untuk mencari referensi.

##### **3.2.2. Kebutuhan Perangkat Keras**

Dalam penelitian ini, dibutuhkan sebuah laptop dengan spesifikasi processor AMD Ryzen 7 5700U, kapasitas SSD 500 GB dan RAM sebesar 16 GB.



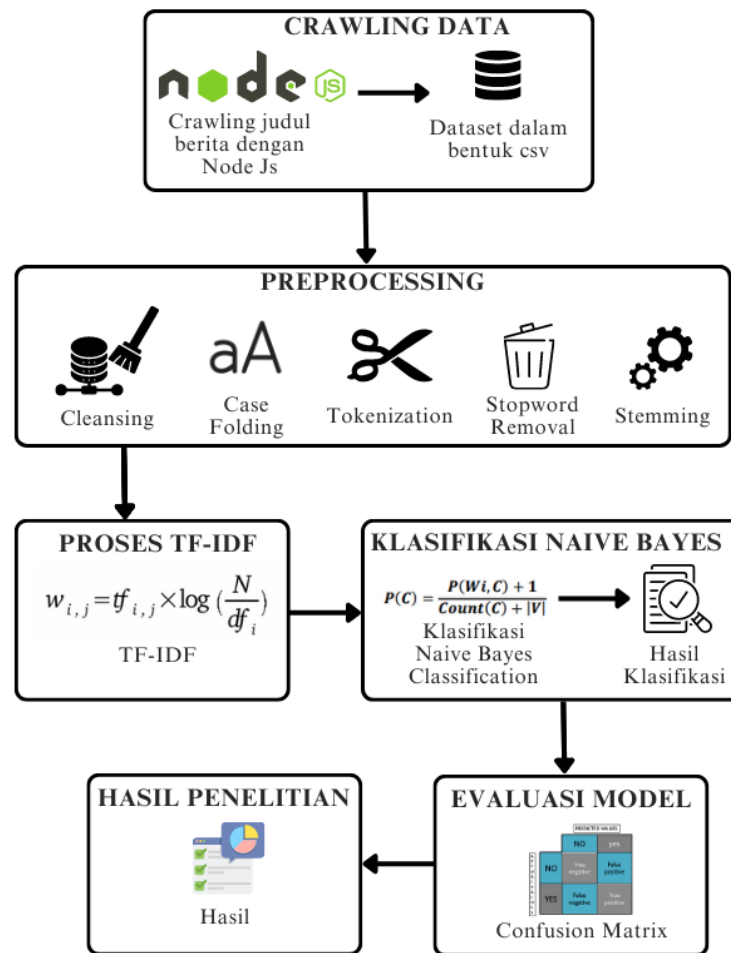
### 3.3. Metode Yang Diusulkan

Alur pelaksanaan metode yang diterapkan dalam penelitian ini diawali dengan melakukan pengumpulan atau *crawling* data judul berita dari website Kompas.TV Kota Semarang, menggunakan bahasa pemrograman Node JS dan memanfaatkan API Kompas.TV. Selanjutnya adalah melakukan pelabelan terhadap data yang sudah didapatkan, sehingga didapatkan data positif dan negatif berdasarkan sentimen yang terkandung dalam judul berita.

Tahap selanjutnya adalah melakukan tahap *preprocessing* data. Tujuan dari tahap *preprocessing* adalah untuk menghilangkan elemen-elemen yang tidak digunakan dalam proses klasifikasi analisis sentimen. Proses perhitungan sentimen dilakukan dengan menggunakan metode *Naïve Bayes Classifier*. Metode ini mengklasifikasikan sentimen berdasarkan judul berita yang telah melalui tahap *preprocessing* dan pembobotan. *Naïve Bayes Classifier* didasarkan pada teorema Bayes dan mengasumsikan independensi antara setiap fitur dalam data.

Setelah data diolah melalui proses *training* dan *testing* menggunakan metode *Naïve Bayes Classifier*, langkah selanjutnya adalah melakukan evaluasi atau pengujian performa. Evaluasi dilakukan dengan memanfaatkan tabel *confusion matrix*, yang digunakan untuk mengetahui tingkat akurasi metode pada proses penelitian yang dilakukan.

Dengan melalui alur pelaksanaan ini, diharapkan metode yang diusulkan dapat menghasilkan klasifikasi sentimen yang akurat berdasarkan judul berita pada media Kompas.TV Kota Semarang. Untuk memudahkan pemahaman terhadap alur pelaksana metode dalam penelitian ini, berikut adalah diagram yang menjelaskan alur pelaksana metode:



Gambar 3.1. Alur Penelitian

### 3.3.1. Crawling Dataset

Pada penelitian ini data dikumpulkan dengan menggunakan cara *crawling*, dimana menggunakan API dari website Kompas.TV yang diolah dibangun menggunakan bahasa pemrograman Node JS. Data yang digunakan masih menggunakan data yang didapat langsung dari *website* Kompas.TV, dengan jumlah data 1886.

### 3.3.2. Preprocessing

*Preprocessing* dalam penelitian ini bertujuan untuk mengorganisir data judul berita agar lebih terstruktur. Selain itu, tahap ini juga dilakukan agar sistem dapat mengenali data dengan lebih mudah saat proses klasifikasi. Tahapan yang harus dilakukan dalam penelitian ini, yaitu:

- a. *Cleansing*, merupakan tahap pembersihan data dilakukan untuk menghilangkan karakter atau informasi yang tidak relevan atau mengganggu dalam judul berita.

Tabel 3.1. Tabel Data Set

No.	Judul Berita
1.	Bantuan Modal Rp 250 Miliar Digelontorkan Bagi UMKM Sembako
2.	Warung Makan di Jalur Pantura Semarang Kendal Terbakar
3.	Gara-gara VCS, Banyak Mahasiswa di Semarang Jadi Korban Pemerasan hingga Rugi Jutaan Rupiah
4.	Mahasiswa PPG Prajabatan UPGRIS Semarang Pamerkan 80 Karya Inovasi
5.	Suasana Haru Pemakaman Korban Kecelakaan Mobil Tertimpa Truk di Semarang

Tabel 3.2. Cleansing

No.	Judul Berita
1.	Bantuan Modal Rp 250 Miliar Digelontorkan Bagi UMKM Sembako
2.	Warung Makan di Jalur Pantura Semarang Kendal Terbakar
3.	Gara gara VCS Banyak Mahasiswa di Semarang Jadi Korban Pemerasan hingga Rugi Jutaan Rupiah
4.	Mahasiswa PPG Prajabatan UPGRIS Semarang Pamerkan 80 Karya Inovasi
5.	Suasana Haru Pemakaman Korban Kecelakaan Mobil Tertimpa Truk di Semarang

- b. *Case Folding*, merupakan tahap dalam *preprocessing* yang dilakukan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil semua.

Tabel 3.3. *Case Folding*

No.	Judul Berita
1.	bantuan modal rp 250 miliar digelontorkan bagi umkm sembako
2.	warung makan di jalur pantura semarang kendal terbakar
3.	gara gara vcs banyak mahasiswa di semarang jadi korban pemerasan hingga rugi jutaan rupiah
4.	mahasiswa ppg prajabatan upgris semarang pameran 80 karya inovasi
5.	suasana haru pemakaman korban kecelakaan mobil tertimpa truk di semarang

- c. *Tokenizing*, merupakan tahapan dalam *preprocessing* yang dilakukan untuk memisahkan teks menjadi unit-unit yang lebih kecil, yang disebut dengan token. Token dapat berupa kata, frasa, atau karakter tertentu yang memiliki makna dalam konteks tertentu. Tujuan dari tokenizing adalah untuk memudahkan pemrosesan dan analisis teks selanjutnya, seperti klasifikasi atau analisis sentimen

Tabel 3.4. *Tokenizing*

No.	Judul Berita
1.	“bantuan”, “modal”, “rp”, “250”, “miliar”, “digelontorkan”, “bagi”, “umkm”, “sembako”
2.	“warung”, “makan”, “di”, “jalur”, “pantura”, “semarang”, “kendal”, “terbakar”
3.	“gara”, “gara”, “vcs”, “banyak”, “mahasiswa”, “di”, “semarang”, “jadi”, “korban”, “pemerasan”, “hingga”, “rugi”, “jutaan”, “rupiah”
4.	“mahasiswa”, “ppg”, “prajabatan”, “upgris”, “semarang”, “pamerkan”, “80”, “karya”, “inovasi”
5.	“suasana”, “haru”, “pemakaman”, “korban”, “kecelakaan”, “mobil”, “tertimpa”, “truk”, “di”, “Semarang”

- d. *Stopword Removal*, merupakan tahap dalam *preprocessing* yang dilakukan untuk menghapus kata-kata yang umum dan tidak memberikan informasi yang berguna dalam analisis teks. Kata-kata tersebut disebut *stopwords*.

Tabel 3.5 Stopword Removal

No.	Judul Berita
1.	bantuan,modal,rp,250,miliar,digelontorkan,bagi,umkm,sembako
2.	warung,jalur,pantura,kendal,terbakar
3.	gara,gara,vcs,jadi,korban,pemerasan,rugi,jutaan,rupiah
4.	ppg,prajabatan,upgris,pamerkan,80,karya,inovasi
5.	suasana,haru,pemakaman,korban,kecelakaan,tertimpa

- e. *Stemming*, tahapan dalam *preprocessing* yang dilakukan untuk mengubah kata-kata menjadi bentuk dasarnya atau kata dasar, dengan menghilangkan kata imbuhan seperti di-, ke-, -nya, me-, ber-, per-, -an, -I, dll dan dilakukan pencocokan kata dasar didalam kamus Bahasa Indonesia

Tabel 3.6. *Stemming*

No.	Judul Berita
1.	bantu,modal,miliar,gelontor,bagi
2.	warung,jalur,kendal,bakar
3.	jadi,korban,peras,rugi,rupiah
4.	pamer,karya,inovasi
5.	suasana,haru,makam,korban,celaka,timpa

### 3.3.3. Pembobotan Kata

Setelah tahap preprocessing selesai, langkah selanjutnya adalah melakukan pembobotan TF-IDF. Pada tahap ini maka terlihat frekuensi

jumlah dari setiap kata serta setiap judul berita diberikan kelas sesuai dengan sentimen.

Tabel 3.7. *Pelabelan sentimen*

No.	Judul Berita	Kategori Kelas
1.	bantu,modal,miliar,gelontor,bagi	Positif
2.	warung,jalur,kendal,bakar	Negatif
3.	jadi,korban,peras,rugi,rupiah	Negatif
4.	pamer,karya,inovasi	Positif
5.	suasana,haru,makam,korban,celaka,timpa	Negatif

#### 3.3.4. Klasifikasi Metode

Langkah selanjutnya setelah dilakukannya proses pembobotan TF-IDF, dilanjutkan dengan melakukan masukan terhadap nilai pada hasil TF-IDF ke persamaan rumus klasifikasi dari algoritma *Naive Bayes*.

#### 3.3.5. Pengujian Metode

Pada penelitian kali ini, metode yang digunakan untuk mengevaluasi kinerja dari algoritma yang digunakan adalah metode *Confussion Matrix*. dalam metode ini, dilakukan perhitungan untuk menentukan nilai TP, TN, FP, FN.

Tabel 3.8 Confussion Matrix

	Actual Values	
Predicted Values	Positive	Negative
Positive	TP	FP
Negative	FN	TN

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Pengumpulan Data

Pada penelitian ini, tahap pertama yang dilakukan adalah mengumpulkan data berupa teks dari judul berita Kompas TV di wilayah Kota Semarang. Proses pengambilan data dilaksanakan dengan menggunakan teknik pengambilan data (*crawling*) yang dikembangkan menggunakan Node JS. Jumlah dataset yang telah dikumpulkan sebanyak 1886 judul berita. Dataset ini kemudian dibagi menjadi dua, yaitu data training sebesar 90% dan data testing sebesar 10%.

Tabel 4.1 Contoh Pelabelan Dataset

NO	Judul Berita	Sentimen
1	Ikhtiar Hadapi Kemarau, Siswa SMK di Semarang Gelar Shalat Minta Hujan	Positive
2	Viral Cuaca Panas Bikin Warga Semarang Manfaatkan Cahaya Matahari untuk Memasak, Ternyata Matang!	Negative
3	Terungkap Isi Surat Mahasiswi Yang Tewas di Kamar Kos	Negative
...	...	...
1884	Banjir di Stasiun Semarang Tawang, Pihak PT KAI Siagakan 3 Pompa Portabel Guna Kurangi Debit Air!	Negative
1885	Stasiun Tawang Semarang Direndam Banjir, Perjalanan Kereta Ditunda!	Negative
1886	Ada 30 Titik Banjir di Semarang, Sebagian Mulai Surut Masuk Hari Pertama Tahun 2023	Positive

#### 4.2. Tahap Preprocessing Data

Tahapan dalam proses penelitian selanjutnya merupakan *proprocessing* data. Pada tahapan ini, bertujuan untuk melakukan normalisasi *dataset*, sehingga data tersebut dapat dianalisa secara maksimal. Pada tahapan ini ada beberapa proses yang dilalui oleh data yang ada, proses tersebut diantaranya *cleansing*, *case folding*,

*tokenizing, stopwords removal, dan stemming.* Berikut adalah tabel hasil dari proses-proses yang dilakukan :

Tabel 4.2.1 Hasil Sampel Dataset *Cleansing*

No	Dokumen
1	Cemburu Suami Nekat Aniaya Istri Hingga Tewas
2	Berkaca pada Kecelakaan Kereta Api Brantas di Semarang Ini Penyebab Kendaraan Mogok di Rel SINAU
3	Sopir Truk Tangki Kecelakaan Maut Jatibarang Jadi Tersangka
4	Edukasi Kesehatan Paguyuban RS Gelar Public Health Fair
5	Kemarau Panjang Ganggu Urban Farming di Kota Semarang
6	Rumah Perakitan dan Jual Beli Senjata Api Ilegal di Semarang
7	Harga Beras di Semarang Naik Rp Per Kilogram
8	Sempat Bertengkar Suami Diduga Aniaya Istri Hingga Tewas
9	Keren Rawat Budaya Lewat Pameran Keris Nusantara di Semarang
10	Krisis Air Bersih di Jawa Tengah BPBD Salurkan Bantuan Air Bersih untuk Warga Terdampak
11	Seru Lomba Estafet Air Hingga Tangkap Bebek Mata Tertutup
12	Menteri Bahlil Optimis Investasi Bakal Naik
13	Lebih Aman Petani Basmi Hama Tikus dengan Gropyokan
14	BPBD Kota Semarang Droping Air Bersih di Empat Kecamatan
15	Menikmati Lobster Rawa Bakar di Rawa Pening
16	Depresi Seorang Pria Nekat Panjat Tower Setinggi Meter
17	Mobil Damkar dan Tangki Air Disiagakan Hingga TPA Jatibarang Tak Berasap
18	Tambang Ilegal di Jateng Marak Harus Segera Ada Legalitas
19	Kemarau Panjang Rawan Kebakaran Penjualan Apar Naik
20	Garuda Nusantara Indonesia Emas Deklarasi Dukung Prabowo di Pilpres
21	Pra Rekonstruksi Kasus Tewasnya Sopir Taksi Online
22	Untuk Meraih Kemenangan Persik Manfaatkan Pemain PSIS yang Absen



23	Petani di Semarang Gotong Royong Tangkap Tikus di Sawah dengan Cara Gropyokan
24	Gaet Investor DPMPTSP Kota Semarang Buat Raperda Insentif Investasi
25	Laka Beruntun di Jatibarang Orang Tewas Orang Luka luka
26	Racik dan Edarkan Kosmetik Ilegal Seorang Pemuda Ditangkap
27	Nikmatnya Soto Sawah Nasi Organik Murah dan Sehat
28	Keren Siswa SD Buat Animasi Edukasi Cegah Banjir dan Rob
29	Sopir Truk Kecelakaan Maut di Semarang Jadi Tersangka
30	Siswa Di Demak Bacok Guru Saat Penilaian Tengah Semester

Tabel 4.2.2 Hasil Sampel Data *Training Case Folding*

No	Dokumen
1	cemburu suami nekat aniaya istri hingga tewas
2	berkaca pada kecelakaan kereta api brantas di semarang ini penyebab kendaraan mogok di rel sinau
3	sopir truk tangki kecelakaan maut jatibarang jadi tersangka
4	edukasi kesehatan paguyuban rs gelar public health fair
5	kemarau panjang ganggu urban farming di kota semarang
6	rumah perakitan dan jual beli senjata api ilegal di semarang
7	harga beras di semarang naik rp per kilogram
8	sempat bertengkar suami diduga aniaya istri hingga tewas
9	keren rawat budaya lewat pameran keris nusantara di semarang
10	krisis air bersih di jawa tengah bpbd salurkan bantuan air bersih untuk warga terdampak
11	seru lomba estafet air hingga tangkap bebek mata tertutup
12	menteri bahlil optimis investasi bakal naik
13	lebih aman petani basmi hama tikus dengan gropyokan
14	bpbd kota semarang droping air bersih di empat kecamatan
15	menikmati lobster rawa bakar di rawa pening

16	depresi seorang pria nekat panjat tower setinggi meter
17	mobil damkar dan tangki air disiagakan hingga tpa jatibarang tak berasap
18	tambang ilegal di jateng marak harus segera ada legalitas
19	kemarau panjang rawan kebakaran penjualan apar naik
20	garuda nusantara indonesia emas deklarasi dukung prabowo di pilpres
21	pra rekonstruksi kasus tewasnya sopir taksi online
22	untuk meraih kemenangan persik memanfaatkan pemain psis yang absen
23	petani di semarang gotong royong tangkap tikus di sawah dengan cara gropyokan
24	gaet investor dpmptsp kota semarang buat raperda insentif investasi
25	laka beruntun di jatibarang orang tewas orang lukaluka
26	racik dan edarkan kosmetik ilegal seorang pemuda ditangkap
27	nikmatnya soto sawah nasi organik murah dan sehat
28	keren siswa sd buat animasi edukasi cegah banjir dan rob
29	sopir truk kecelakaan maut di semarang jadi tersangka
30	siswa di demak bacok guru saat penilaian tengah semester

Tabel 4.2.3 Hasil Sampel Dataset *Tokenizing*

No	Dokumen
1	'cemburu', 'suami', 'nekat', 'aniaya', 'istri', 'hingga', 'tewas'
2	'berkaca', 'pada', 'kecelakaan', 'kereta', 'api', 'brantas', 'di', 'semarang', 'ini', 'penyebab', 'kendaraan', 'mogok', 'di', 'rel', 'sinau'
3	'sopir', 'truk', 'tangki', 'kecelakaan', 'maut', 'jatibarang', 'jadi', 'tersangka'
4	'edukasi', 'kesehatan', 'paguyuban', 'rs', 'gelar', 'public', 'health', 'fair'
5	'kemarau', 'panjang', 'ganggu', 'urban', 'farming', 'di', 'kota', 'semarang'
6	'rumah', 'perakitan', 'dan', 'jual', 'beli', 'senjata', 'api', 'ilegal', 'di', 'semarang'
7	'harga', 'beras', 'di', 'semarang', 'naik', 'rp', '2', '500', 'per', 'kilogram'
8	'sempat', 'bertengkar', 'suami', 'diduga', 'aniaya', 'istri', 'hingga', 'tewas'

9	'keren', 'rawat', 'budaya', 'lewat', 'pameran', 'keris', 'nusantara', 'di', 'semarang'
10	'krisis', 'air', 'bersih', 'di', 'jawa', 'tengah', 'bpbd', 'salurkan', 'bantuan', 'air', 'bersih', 'untuk', 'warga', 'terdampak'
11	'seru', 'lomba', 'estafet', 'air', 'hingga', 'tangkap', 'bebek', 'mata', 'tertutup'
12	'menteri', 'bahlil', 'optimis', 'investasi', '2023', 'bakal', 'naik'
13	'lebih', 'aman', 'petani', 'basmi', 'hama', 'tikus', 'dengan', 'gropyokan'
14	'bpbd', 'kota', 'semarang', 'droping', 'air', 'bersih', 'di', 'empat', 'kecamatan'
15	'menikmati', 'lobster', 'rawa', 'bakar', 'di', 'rawa', 'pening'
16	'depresi', 'seorang', 'pria', 'nekat', 'panjat', 'tower', 'setinggi', '50', 'meter'
17	'mobil', 'damkar', 'dan', 'tangki', 'air', 'disiagakan', 'hingga', 'tpa', 'jatibarang', 'tak', 'berasap'
18	'tambang', 'ilegal', 'di', 'jateng', 'marak', 'harus', 'segera', 'ada', 'legalitas'
19	'kemarau', 'panjang', 'rawan', 'kebakaran', 'penjualan', 'apar', 'naik'
20	'garuda', 'nusantara', '08', 'indonesia', 'emas', 'deklarasi', 'dukung', 'prabowo', 'di', 'pilpres', '2024'
21	'pra', 'rekonstruksi', 'kasus', 'tewasnya', 'sopir', 'taksi', 'online'
22	'untuk', 'meraih', 'kemenangan', 'persik', 'manfaatkan', 'pemain', 'psis', 'yang', 'absen'
23	'petani', 'di', 'semarang', 'gotong', 'royong', 'tangkap', 'tikus', 'di', 'sawah', 'dengan', 'cara', 'gropyokan'
24	'gaet', 'investor', 'dpmpst', 'kota', 'semarang', 'buat', 'raperda', 'insentif', 'investasi'
25	'laka', 'beruntun', 'di', 'jatibarang', '1', 'orang', 'tewas', '3', 'orang', 'luka', 'luka'
26	'racik', 'dan', 'edarkan', 'kosmetik', 'ilegal', 'seorang', 'pemuda', 'ditangkap'
27	'nikmatnya', 'soto', 'sawah', 'nasi', 'organik', 'murah', 'dan', 'sehat'
28	'keren', 'siswa', 'sd', 'buat', 'animasi', 'edukasi', 'cegah', 'banjir', 'dan', 'rob'
29	'sopir', 'truk', 'kecelakaan', 'maut', 'di', 'semarang', 'jadi', 'tersangka'
30	'siswa', 'di', 'demak', 'bacok', 'guru', 'saat', 'penilaian', 'tengah', 'semester'

Tabel 4.2.4 Hasil Sampel Dataset *Stopword Removal*

No	Dokumen
1	cemburu,suami,nekat,aniaya,istri,tewas
2	berkaca,kecelakaan,api,brantas,penyebab,kendaraan,mogok,rel,sinau
3	tangki,kecelakaan,maut,jatibarang,jadi,tersangka
4	edukasi,kesehatan,paguyuban,rs,gelar,public,health,fair
5	kemarau,ganggu,farming
6	perakitan,jual,senjata,api,ilegal
7	harga,naik,rp,2,500
8	bertengkar,suami,diduga,aniaya,istri,tewas
9	keren,rawat,budaya,pameran,keris
10	krisis,bersih,jawa,bpbd,salurkan,bantuan,bersih,terdampak
11	seru,estafet,tangkap,mata,tertutup
12	bahlil,optimis,investasi,2023,naik
13	aman,petani,basmi,hama,tikus,gropyokan
14	bpbd,droping,bersih,kecamatan
15	menikmati,lobster,rawa,bakar,rawa,pening
16	depresi,pria,nekat,tower,50
17	damkar,tangki,disiagakan,tpa,jatibarang,berasap
18	tambang,ilegal,jateng,marak,harus,legalitas
19	kemarau,rawan,kebakaran,penjualan,apar,naik
20	08,emas,dukung,prabowo,pilpres,2024
21	pra,rekonstruksi,kasus,tewasnya,taksi,online
22	meraih,kemenangan,persik,manfaatkan,pemain,psis,absen
23	petani,gotong,royong,tangkap,tikus,sawah,gropyokan
24	gaet,investor,dpmptsp,buat,raperda,insentif,investasi
25	laka,beruntun,jatibarang,1,tewas,3,luka,luka
26	racik,edarkan,kosmetik,ilegal,ditangkap
27	nikmatnya,soto,sawah,organik,murah,sehat

28	keren,sd,buat,animasi,edukasi,cegah,banjir,rob
29	kecelakaan,maut,jadi,tersangka
30	siswa,bacok,guru ,tengah,semester

Tabel 4.2.5 Hasil Sampel Dataset *Stemming*

No	Dokumen
1	cemburu,suami,nekat,aniaya,istri,tewas
2	kaca,celaka,api,kendara,mogok,rel,sinau
3	tangki,celaka,maut,jadi,sangka
4	edukasi,sehat,paguyuban,gelar
5	kemarau,ganggu
6	rakit,jual,senjata,api,ilegal
7	harga,naik
8	tengkar,suami,duga,aniaya,istri,tewas
9	keren,rawat,budaya,pamer,keris
10	krisis,bersih,salur,bantu,bersih,dampak
11	seru,estafet,tangkap,mata,tutup
12	optimis,investasi,naik
13	aman,tani,basmi,hama,tikus
14	bersih,camat
15	nikmat,lobster,rawa,bakar,rawa,pening
16	depresi,pria,nekat
17	tangki,siaga,asap
18	tambang,ilegal,marak,harus,legalitas
19	kemarau,rawan,bakar,jual,apar,naik
20	emas,dukung
21	rekonstruksi,kasus,tewas,taksi
22	raih,menang,persik,manfaat,absen
23	tani,gotong,tangkap,tikus,sawah

24	gaet,investor,buat,insentif,investasi
25	tewas,luka,luka
26	racik,edar,kosmetik,ilegal,tangkap
27	nikmat,soto,sawah,organik,murah,sehat
28	keren,buat,animasi,edukasi,cegah,banjir
29	celaka,maut,jadi,sangka
30	bacok,guru,semester

#### 4.3. Pembobotan Kata

Tahap selanjutnya adalah proses TF-IDF, dataset akan dilakukan proses pembobotan kata. Berikut merupakan dataset sampel yang sudah melalui proses pembobotan kata:

Tabel 4.3.1 Tabel Perhitungan TF-IDF Data *Training*

Term	TF	DF	IDF	TF-IDF	Sentimen
cemburu	1	2	1.454845	1.454845	Negative
suami	1	4	1.153815	1.153815	Negative
nekat	1	4	1.153815	1.153815	Negative
aniaya	1	4	1.153815	1.153815	Negative
istri	1	4	1.153815	1.153815	Negative
tewas	1	8	0.852785	0.852785	Negative
kaca	1	2	1.454845	1.454845	Negative
celaka	1	5	1.056905	1.056905	Negative
api	1	4	1.153815	1.153815	Negative
kendara	1	2	1.454845	1.454845	Negative
mogok	1	2	1.454845	1.454845	Negative
rel	1	2	1.454845	1.454845	Negative
sinau	1	2	1.454845	1.454845	Negative
tangki	1	4	1.153815	1.153815	Negative
maut	1	3	1.278754	1.278754	Negative
jadi	1	3	1.278754	1.278754	Negative
sangka	1	3	1.278754	1.278754	Negative
edukasi	1	3	1.278754	1.278754	Positive
sehat	1	4	1.153815	1.153815	Positive

paguyuban	1	2	1.454845	1.454845	Positive
gelar	1	2	1.454845	1.454845	Positive
kemarau	1	4	1.153815	1.153815	Negative
ganggu	1	2	1.454845	1.454845	Negative
rakit	1	2	1.454845	1.454845	Negative
jual	1	4	1.153815	1.153815	Negative
senjata	1	2	1.454845	1.454845	Negative
ilegal	1	6	0.977724	0.977724	Negative
harga	1	2	1.454845	1.454845	Negative
naik	1	6	0.977724	0.977724	Negative
tengkar	1	2	1.454845	1.454845	Negative
duga	1	2	1.454845	1.454845	Negative
keren	1	3	1.278754	1.278754	Positive
rawat	1	2	1.454845	1.454845	Positive
budaya	1	2	1.454845	1.454845	Positive
pamer	1	2	1.454845	1.454845	Positive
keris	1	2	1.454845	1.454845	Positive
krisis	1	2	1.454845	1.454845	Negative
bersih	2	4	1.153815	2.30763	Negative
salur	1	2	1.454845	1.454845	Negative
bantu	1	2	1.454845	1.454845	Negative
dampak	1	2	1.454845	1.454845	Negative
seru	1	2	1.454845	1.454845	Positive
estafet	1	2	1.454845	1.454845	Positive
tangkap	1	6	0.977724	0.977724	Positive
mata	1	2	1.454845	1.454845	Positive
tutup	1	2	1.454845	1.454845	Positive
optimis	1	2	1.454845	1.454845	Positive
investasi	1	4	1.153815	1.153815	Positive
aman	1	2	1.454845	1.454845	Positive
tani	1	4	1.153815	1.153815	Positive
basmi	1	2	1.454845	1.454845	Positive
hama	1	2	1.454845	1.454845	Positive
tikus	1	4	1.153815	1.153815	Positive
camat	1	2	1.454845	1.454845	Positive
nikmat	1	4	1.153815	1.153815	Positive
lobster	1	2	1.454845	1.454845	Positive
rawa	2	2	1.454845	2.90969	Positive
bakar	1	4	1.153815	1.153815	Positive
pening	1	2	1.454845	1.454845	Positive
depresi	1	2	1.454845	1.454845	Negative

pria	1	2	1.454845	1.454845	Negative
siaga	1	2	1.454845	1.454845	Negative
asap	1	2	1.454845	1.454845	Negative
tambang	1	2	1.454845	1.454845	Negative
marak	1	2	1.454845	1.454845	Negative
harus	1	2	1.454845	1.454845	Negative
legalitas	1	2	1.454845	1.454845	Negative
rawan	1	2	1.454845	1.454845	Negative
apar	1	2	1.454845	1.454845	Negative
emas	1	2	1.454845	1.454845	Positive
dukung	1	2	1.454845	1.454845	Positive
rekonstruksi	1	2	1.454845	1.454845	Negative
kasus	1	2	1.454845	1.454845	Negative
taksi	1	2	1.454845	1.454845	Negative
raih	1	2	1.454845	1.454845	Positive
menang	1	2	1.454845	1.454845	Positive
persik	1	2	1.454845	1.454845	Positive
manfaat	1	2	1.454845	1.454845	Positive
absen	1	2	1.454845	1.454845	Positive
gotong	1	2	1.454845	1.454845	Positive
sawah	1	4	1.153815	1.153815	Positive
gaet	1	2	1.454845	1.454845	Positive
investor	1	2	1.454845	1.454845	Positive
buat	1	3	1.278754	1.278754	Positive
insentif	1	2	1.454845	1.454845	Positive
luka	2	2	1.454845	2.90969	Negative
racik	1	2	1.454845	1.454845	Negative
edar	1	2	1.454845	1.454845	Negative
kosmetik	1	2	1.454845	1.454845	Negative
soto	1	2	1.454845	1.454845	Positive
organik	1	2	1.454845	1.454845	Positive
murah	1	2	1.454845	1.454845	Positive

Tabel 4.3.2 Tabel Perhitungan TF-IDF Data *Testing*

Term	TF	DF	IDF	TF-IDF
cemburu	1	2	1.454845	1.454845
suami	1	4	1.153815	1.153815
nekat	1	4	1.153815	1.153815
aniaya	1	4	1.153815	1.153815
istri	1	4	1.153815	1.153815



tewas	1	8	0.852785	0.852785
kaca	1	2	1.454845	1.454845
celaka	1	5	1.056905	1.056905
api	1	4	1.153815	1.153815
kendara	1	2	1.454845	1.454845
mogok	1	2	1.454845	1.454845
rel	1	2	1.454845	1.454845
sinau	1	2	1.454845	1.454845
tangki	1	4	1.153815	1.153815
maut	1	3	1.278754	1.278754
jadi	1	3	1.278754	1.278754
sangka	1	3	1.278754	1.278754
edukasi	1	3	1.278754	1.278754
sehat	1	4	1.153815	1.153815
paguyuban	1	2	1.454845	1.454845
gelar	1	2	1.454845	1.454845
kemarau	1	4	1.153815	1.153815
ganggu	1	2	1.454845	1.454845
rakit	1	2	1.454845	1.454845
jual	1	4	1.153815	1.153815
senjata	1	2	1.454845	1.454845
ilegal	1	6	0.977724	0.977724
harga	1	2	1.454845	1.454845
naik	1	6	0.977724	0.977724
tengkar	1	2	1.454845	1.454845
duga	1	2	1.454845	1.454845
keren	1	3	1.278754	1.278754
rawat	1	2	1.454845	1.454845
budaya	1	2	1.454845	1.454845
pamer	1	2	1.454845	1.454845
keris	1	2	1.454845	1.454845
krisis	1	2	1.454845	1.454845
bersih	2	4	1.153815	2.30763
salur	1	2	1.454845	1.454845
bantu	1	2	1.454845	1.454845
dampak	1	2	1.454845	1.454845
seru	1	2	1.454845	1.454845
estafet	1	2	1.454845	1.454845
tangkap	1	6	0.977724	0.977724
mata	1	2	1.454845	1.454845
tutup	1	2	1.454845	1.454845
optimis	1	2	1.454845	1.454845

investasi	1	4	1.153815	1.153815
aman	1	2	1.454845	1.454845
tani	1	4	1.153815	1.153815
basmi	1	2	1.454845	1.454845
hama	1	2	1.454845	1.454845
tikus	1	4	1.153815	1.153815
camat	1	2	1.454845	1.454845
nikmat	1	4	1.153815	1.153815
lobster	1	2	1.454845	1.454845
rawa	2	2	1.454845	2.90969
bakar	1	4	1.153815	1.153815
pening	1	2	1.454845	1.454845
depresi	1	2	1.454845	1.454845
pria	1	2	1.454845	1.454845
siaga	1	2	1.454845	1.454845
asap	1	2	1.454845	1.454845
tambang	1	2	1.454845	1.454845
marak	1	2	1.454845	1.454845
harus	1	2	1.454845	1.454845
legalitas	1	2	1.454845	1.454845
rawan	1	2	1.454845	1.454845
apar	1	2	1.454845	1.454845
emas	1	2	1.454845	1.454845
dukung	1	2	1.454845	1.454845
rekonstruksi	1	2	1.454845	1.454845
kasus	1	2	1.454845	1.454845
taksi	1	2	1.454845	1.454845
raih	1	2	1.454845	1.454845
menang	1	2	1.454845	1.454845
persik	1	2	1.454845	1.454845
manfaat	1	2	1.454845	1.454845
absen	1	2	1.454845	1.454845
gotong	1	2	1.454845	1.454845
sawah	1	4	1.153815	1.153815
gaet	1	2	1.454845	1.454845
investor	1	2	1.454845	1.454845
buat	1	3	1.278754	1.278754
insentif	1	2	1.454845	1.454845
luka	2	2	1.454845	2.90969
racik	1	2	1.454845	1.454845
edar	1	2	1.454845	1.454845
kosmetik	1	2	1.454845	1.454845

soto	1	2	1.454845	1.454845
organik	1	2	1.454845	1.454845
murah	1	2	1.454845	1.454845
animasi	1	1	1.755875	1.755875
cegah	1	1	1.755875	1.755875
banjir	1	1	1.755875	1.755875
bacok	1	1	1.755875	1.755875
guru	1	1	1.755875	1.755875
semester	1	1	1.755875	1.755875

#### 4.4. Klasifikasi Metode

Tahapan selanjutnya merupakan pembagian dataset menjadi dua kelas, yaitu data *training* dan data *testing*. Berikut merupakan rumus mencari probabilitas Naïve Bayes:

$$P(C) = \frac{P(W_i, C) + 1}{\text{Count}(C) + |V|}$$

1. Dokumen ke-1 (cemburu, suami, nekat, aniaya, istri, tewas)

$$P(\text{cemburu}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{cemburu}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{suami}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{suami}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{nekat}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{nekat}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{aniaya}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(aniaya|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(istri|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(istri|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(tewas|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(tewas|Negatif) = \frac{4 + 1}{86.61476 + 98} = 0.05773$$

$$Total P( Positif ) = 0.08459$$

$$Total P( Negatif ) = 0.21936$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-1 bersentimen negatif

2. Dokumen ke-2 (kaca, celaka, api, kendaraan, mogok, rel, sinau)

$$P(kaca|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(kaca|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(celaka|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(celaka|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(api|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(api|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(kendaraan|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(kendara|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(mogok|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(mogok|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(rel|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(rel|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(sinau|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(sinau|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.09869$$

$$Total P( Negatif ) = 0.18473$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-2 bersentimen negatif

3. Dokumen ke-3 (tangki, celaka, maut, jadi, sangka)

$$P(celaka|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(celaka|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(tangki|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(tangki|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(maut|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(maut|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(jadi|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(jadi|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(sangka|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(sangka|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.07049$$

$$Total P( Negatif ) = 0.13854$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-3 bersentimen negatif

4. Dokumen ke-4 (edukasi, sehat, paguyuban, gelar)

$$P(edukasi|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(edukasi|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(sehat|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(sehat|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(paguyuban|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(paguyuban|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(gelar|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(gelar|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$Total P( Positif ) = 0.12688$$

$$Total P( Negatif ) = 0.04618$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-4 bersentimen positif

5. Dokumen ke-5 (kemarau, ganggu)

$$P(kemarau|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(kemarau|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(ganggu|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(ganggu|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.02820$$

$$Total P( Negatif ) = 0.05773$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-5 bersentimen negatif

6. Dokumen ke-6 (rakit, jual, senjata, api, ilegal)

$$P(api|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(api|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(rakit|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(rakit|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(jual|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(jual|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(senjata|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(senjata|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(ilegal|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(ilegal|Negatif) = \frac{3 + 1}{86.61476 + 98} = 0.04618$$

$$Total P( Positif ) = 0.07049$$

$$Total P( Negatif ) = 0.16164$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-6 bersentimen negatif

7. Dokumen ke-7 (harga, naik)

$$P(harga|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(harga|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(naik|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(naik|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$Total P( Positif ) = 0.04229$$

$$Total P( Negatif ) = 0.05773$$



Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-7 bersentimen negatif

8. Dokumen ke-8 (tengkar, suami, duga, aniaya, istri, tewas)

$$P(\text{suami}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{suami}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{aniaya}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{aniaya}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{istri}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{istri}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{tewas}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{tewas}|\text{Negatif}) = \frac{4 + 1}{86.61476 + 98} = 0.05773$$

$$P(\text{tengkar}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{tengkar}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{duga}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{duga}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$\text{Total } P(\text{Positif}) = 0.08459$$

$$\text{Total } P(\text{Negatif}) = 0.20782$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-8 bersentimen negatif

9. Dokumen ke-9 (keren, rawat, budaya, pamer, keris)

$$P(keren|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(keren|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(rawat|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(rawat|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(budaya|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(budaya|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(pamer|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(pamer|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(keris|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(keris|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$Total P( Positif ) = 0.14098$$

$$Total P( Negatif ) = 0.05773$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-9 bersentimen positif

10. Dokumen ke-10 (krisis, bersih, salur, bantu, bersih, dampak)

$$P(krisis|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(krisis|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(bersih|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(bersih|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(salur|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(salur|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(bantu|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(bantu|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(dampak|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(dampak|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.08459$$

$$Total P( Negatif ) = 0.11545$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-10 bersentimen negatif

11. Dokumen ke-11 (seru, estafet, tangkap, mata, tutup)

$$P(seru|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{seru}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{estafet}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{estafet}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{tangkap}|\text{Positif}) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(\text{tangkap}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{mata}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{mata}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{tutup}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{tutup}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.15508$$

$$\text{Total } P(\text{Negatif}) = 0.06927$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-11 bersentimen positif

12. Dokumen ke-12 (optimis, investasi, naik)

$$P(\text{naik}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{naik}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{optimis}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{optimis}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{investasi}|\text{Positif}) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(\text{investasi}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.09869$$

$$\text{Total } P(\text{Negatif}) = 0.05773$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-12 bersentimen positif

13. Dokumen ke-13 (aman, tani, basmi, hama, tikus)

$$P(\text{aman}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{aman}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{tani}|\text{Positif}) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(\text{tani}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{basm}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{basm}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{hama}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{hama}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{tikus}|\text{Positif}) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(tikus|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$Total P( Positif ) = 0.16917$$

$$Total P( Negatif ) = 0.05773$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-13 bersentimen positif

14. Dokumen ke-14 (bersih, camat)

$$P(bersih|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(bersih|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(camat|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(camat|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$Total P( Positif ) = 0.05639$$

$$Total P( Negatif ) = 0.03464$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-14 bersentimen positif

15. Dokumen ke-15 (nikmat, lobster, rawa, bakar, rawa, pening)

$$P(nikmat|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(nikmat|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(lobster|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(lobster|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{rawa}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{rawa}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{bakar}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{bakar}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{pening}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{pening}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.15508$$

$$\text{Total } P(\text{Negatif}) = 0.06927$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-15 bersentimen positif

16. Dokumen ke-16 (depresi, pria, nekat)

$$P(\text{nekat}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{nekat}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{depresi}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{depresi}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{pria}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{pria}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.04229$$

$$Total P( Negatif ) = 0.08082$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-16 bersentimen negatif

17. Dokumen ke-17 (tangki, siaga, asap)

$$P(tangki|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(tangki|Negatif) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(siaga|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(siaga|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(asap|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(asap|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.04229$$

$$Total P( Negatif ) = 0.08082$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-17 bersentimen negatif

18. Dokumen ke-18 (tambang, ilegal, marak, harus, legalitas)

$$P(ilegal|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(ilegal|Negatif) = \frac{3 + 1}{86.61476 + 98} = 0.04618$$

$$P(tambang|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$



$$P(\text{tambang}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{marak}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{marak}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{harus}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{harus}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{legalitas}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{legalitas}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$\text{Total } P(\text{Positif}) = 0.07049$$

$$\text{Total } P(\text{Negatif}) = 0.13854$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-18 bersentimen negatif

19. Dokumen ke-19 (kemarau, rawan, bakar, jual, apar, naik)

$$P(\text{kemarau}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{kemarau}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{jual}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{jual}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{naik}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{naik}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{bakar}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{bakar}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{rawan}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{rawan}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{apar}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{apar}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$\text{Total } P(\text{Positif}) = 0.11278$$

$$\text{Total } P(\text{Negatif}) = 0.17318$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-19 bersentimen negatif

20. Dokumen ke-20 (emas, dukung)

$$P(\text{emas}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{emas}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{dukung}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{dukung}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.05639$$

$$Total P( Negatif ) = 0.02309$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-20 bersentimen positif

21. Dokumen ke-21 (rekonstruksi, kasus, tewas, taksi)

$$P(tewas|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(tewas|Negatif) = \frac{4 + 1}{86.61476 + 98} = 0.05773$$

$$P(rekonstruksi|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(rekonstruksi|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(kasus|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(kasus|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(taksi|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(taksi|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.05639$$

$$Total P( Negatif ) = 0.12700$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-21 bersentimen negatif

22. Dokumen ke-22 (raih, menang, persik, manfaat, absen)

$$P(raih|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{raih}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{menang}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{menang}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{persik}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{persik}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{manfaat}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{manfaat}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{absen}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{absen}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.14098$$

$$\text{Total } P(\text{Negatif}) = 0.05773$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-22 bersentimen positif

23. Dokumen ke-23 (tani, gotong, tangkap, tikus, sawah)

$$P(\text{tangkap}|\text{Positif}) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(\text{tangkap}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{tani}|\text{Positif}) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(tani|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(tikus|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(tikus|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(gotong|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(gotong|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(sawah|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(sawah|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$Total P( Positif ) = 0.19737$$

$$Total P( Negatif ) = 0.06927$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-23 bersentimen positif

24. Dokumen ke-24 (gaet, investor, buat, insentif, investasi)

$$P(investasi|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(investasi|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(gaet|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(gaet|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(investor|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{investor}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{buat}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{buat}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{insentif}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{insentif}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.15508$$

$$\text{Total } P(\text{Negatif}) = 0.05773$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-24 bersentimen positif

25. Dokumen ke-25 (tewas, luka, luka)

$$P(\text{tewas}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{tewas}|\text{Negatif}) = \frac{4 + 1}{86.61476 + 98} = 0.05773$$

$$P(\text{luka}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{luka}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$\text{Total } P(\text{Positif}) = 0.02820$$

$$\text{Total } P(\text{Negatif}) = 0.08082$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-25 bersentimen negatif

26. Dokumen ke-26 (racik, edar, kosmetik, ilegal, tangkap)

$$P(ilegal|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(ilegal|Negatif) = \frac{3 + 1}{86.61476 + 98} = 0.04618$$

$$P(tangkap|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(tangkap|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(racik|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(racik|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(edar|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(edar|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(kosmetik|Positif) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(kosmetik|Negatif) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$Total P( Positif ) = 0.09869$$

$$Total P( Negatif ) = 0.13854$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-26 bersentimen negatif

27. Dokumen ke-27 (nikmat, soto, sawah, organik, murah, sehat)

$$P(sehat|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(sehat|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(nikmat|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(nikmat|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(sawah|Positif) = \frac{2 + 1}{70.93272 + 98} = 0.04229$$

$$P(sawah|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(soto|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(soto|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(organik|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(organik|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(murah|Positif) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(murah|Negatif) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$Total P( Positif ) = 0.21147$$

$$Total P( Negatif ) = 0.06927$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-27 bersentimen positif



28. Dokumen ke-28 (keren, buat, animasi, edukasi, cegah, banjir)

$$P(\text{edukasi}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{edukasi}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{keren}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{keren}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{buat}|\text{Positif}) = \frac{1 + 1}{70.93272 + 98} = 0.02820$$

$$P(\text{buat}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{animasi}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{animasi}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{cegah}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{cegah}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{banjir}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{banjir}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.12688$$

$$\text{Total } P(\text{Negatif}) = 0.06927$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-28 bersentimen positif

29. Dokumen ke-29 (celaka, maut, jadi, sangka)

$$P(\text{celaka}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{celaka}|\text{Negatif}) = \frac{2 + 1}{86.61476 + 98} = 0.03464$$

$$P(\text{maut}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{maut}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{jadi}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{jadi}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$P(\text{sangka}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{sangka}|\text{Negatif}) = \frac{1 + 1}{86.61476 + 98} = 0.02309$$

$$\text{Total } P(\text{Positif}) = 0.05639$$

$$\text{Total } P(\text{Negatif}) = 0.10391$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-29 bersentimen negatif

30. Dokumen ke-30 (bacok, guru, semester)

$$P(\text{bacok}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{bacok}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{guru}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{guru}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$P(\text{semester}|\text{Positif}) = \frac{0 + 1}{70.93272 + 98} = 0.01410$$

$$P(\text{semester}|\text{Negatif}) = \frac{0 + 1}{86.61476 + 98} = 0.01155$$

$$\text{Total } P(\text{Positif}) = 0.04229$$

$$\text{Total } P(\text{Negatif}) = 0.03464$$

Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-30 bersentimen positif

Tabel 4.4.1 Hasil Prediksi Naïve Bayes

No	Dokumen	Probabilitas Positif	Probabilitas Negatif	Prediksi Sentimen	Sentimen Asli
1	cemburu, suami, nekat, aniaya, istri, tewas	0.08459	0.21936	Negative	Negative
2	kaca, celaka, api, kendaraan, mogok, rel, sianu	0.09869	0.18473	Negative	Negative
3	tangki, celaka, maut, jadi, sangka	0.07049	0.13854	Negative	Negative
4	edukasi, sehat, paguyuban, gelar	0.12688	0.04618	Positive	Positive
5	kemarau, ganggu	0.0282	0.05773	Negative	Negative
6	rakit, jual, senjata, api, ilegal	0.07049	0.16164	Negative	Negative
7	harga, naik	0.04229	0.05773	Negative	Negative
8	tengkar, suami, duga, aniaya, istri, tewas	0.08459	0.20782	Negative	Negative
9	keren, rawat, budaya, pamer, keris	0.14098	0.05773	Positive	Positive
10	krisis, bersih, salur, bantu, bersih, dampak	0.08459	0.11545	Negative	Negative

11	seru, estafet, tangkap, mata, tutup	0.15508	0.06927	Positive	Positive
12	optimis, investasi, naik	0.09869	0.05773	Positive	Positive
13	aman, tani, basmi, hama, tikus	0.16917	0.05773	Positive	Positive
14	bersih, camat	0.05639	0.03464	Positive	Positive
15	nikmat, lobster, rawa, bakar, rawa, pening	0.15508	0.06927	Positive	Positive
16	depresi, pria, nekat	0.04229	0.08082	Negative	Negative
17	tangki, siaga, asap	0.04229	0.08082	Negative	Negative
18	tambang, ilegal, marak, harus, legalitas	0.07049	0.13854	Negative	Negative
19	kemarau, rawan, bakar, jual, apar, naik	0.11278	0.17318	Negative	Negative
20	emas, dukung	0.05639	0.02309	Positive	Positive
21	rekonstruksi, kasus, tewas, taksi	0.05639	0.127	Negative	Negative
22	raih, menang, persik, manfaat, absen	0.14098	0.05773	Positive	Positive
23	tani, gotong, tangkap, tikus, sawah	0.19737	0.06927	Positive	Positive
24	gaet, investor, buat, insentif, investasi	0.15508	0.05773	Positive	Positive
25	tewas, luka, luka	0.0282	0.08082	Negative	Negative
26	racik, edar, kosmetik, ilegal, tangkap	0.09869	0.13854	Negative	Negative
27	nikmat, soto, sawah, organik, murah, sehat	0.21147	0.06927	Positive	Positive
28	keren, buat, animasi, edukasi, cegah, banjir	0.12688	0.06927	Positive	Positive
29	celaka, maut, jadi, sangka	0.05639	0.10391	Negative	Negative
30	bacok, guru, semester	0.04229	0.03464	Positive	Negative

#### 4.5. Pengujian Metode

Dataset dilakukan pembagian menjadi dua bagian, yaitu data latih dan data uji. Perbandingan rasio persentase data latih dan data uji sebesar 90%:10%. Perbandingan ini didapatkan dari evaluasi pengujian yang memiliki nilai tingkat

akurasi tertinggi, dan semakin banyak data uji maka tingkat akurasinya semakin besar. Berikut adalah hasil pengujian metode menggunakan confusion matrix:

Tabel 4.5.1 Confussion Matrix

Predicted Values	Actual Values	
	Positive	Negative
Positive	13	1
Negative	0	16

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FN+FP} * 100\%$$

$$\text{Akurasi} = \frac{13+16}{13+16+0+0} * 100\%$$

$$\text{Akurasi} = 96.67\%$$

$$\text{Presisi kelas negatif} = \frac{TN}{TN+FN} = \frac{16}{16+0} = 1$$

$$\text{Recall kelas negatif} = \frac{TN}{TN+FP} = \frac{16}{16+1} = 0.94$$

$$\text{Presisi kelas positif} = \frac{TP}{TP+FP} = \frac{13}{13+1} = 0.93$$

$$\text{Recall kelas positif} = \frac{TP}{TP+FN} = \frac{13}{13+0} = 1$$

Dari data diatas diperoleh nilai akurasi sebesar 96.67%, nilai presisi kelas negatif sebesar 0.93 dan *recall* kelas negatif sebesar 1 dengan menggunakan 90% data latih dan 10% data uji.

#### 4.6. Hasil Evaluasi

Setelah dilakukan klasifikasi, langkah selanjutnya yaitu mengevaluasi metode yang digunakan. Dalam evaluasi metode ini, data dalam penelitian ini dibagi menjadi 90% data latih dan 10% data uji.

Tabel 4.6.1 Perbandingan Hasil Performa NBC Rasio Dataset

Rasio	Data Latih	Data Uji	Akurasi	Presisi Negatif	Recall Negatif	Presisi Positif	Recall Positif
90:10	27	3	96.67%	100%	94%	93%	100%
85:15	26	4	96.67%	100%	94%	93%	100%
80:10	24	6	93.33%	100%	88%	87%	100%

Hasil dari proses evaluasi metode menggunakan confusion matrix dapat diketahui dari melalui nilai presisi serta *recall* pada setiap kelasnya. Pada penelitian ini, nilai presisi atau tingkat kemampuan sistem untuk mencari ketepatan informasi yang diminta sebesar 93% pada kelas positif, sedangkan kelas negatif sebesar 100%. Sedangkan tingkat keberhasilan sistem untuk menemukan kembali informasi atau *recall* pada kelas positif sebesar 100% dan pada kelas negatif sebesar 94%. Sehingga penelitian ini mendapatkan nilai akurasi sebesar 96.67%.

#### 4.7. Pseudocode Program

Dalam rangkaian penelitian ini, terlampir di bawah adalah algoritma pseudocode yang secara rinci menguraikan program yang telah dikembangkan sebelumnya:

##### 4.7.1. Inisialisasi import library

```

1 Inisialisasi:
2   - Import library readline-sync
3   - Import library figlet
4   - Import library chalk
5   - Import library axios
6   - Import library jsdom
7   - Import JSDOM dari jsdom
8   - Import library fs
9   - Import library natural
10  - Buat objek tokenizer dari natural.WordTokenizer
11  - Import sastrawijs
12  - Buat objek stemmer dari sastrawi.Stemmer

```

Gambar 4.1. Pseudocode Inisialisasi Import Library

### 4.7.2. Inisialisasi variabel global

```

1 - Inisialisasi Variabel:
2   - Inisialisasikan variabel o dengan nilai 1
3   - Inisialisasikan array Judul sebagai wadah untuk menyimpan judul
4   - Inisialisasikan variabel trainingData sebagai wadah untuk data pelatihan
5   - Inisialisasikan variabel testData sebagai wadah untuk data uji
6   - Inisialisasikan variabel combinedData sebagai wadah untuk menggabungkan data pelatihan dan data uji
7   - Inisialisasikan variabel positiveSentimentTFIDF untuk menyimpan hasil TF-IDF positif
8   - Inisialisasikan variabel negativeSentimentTFIDF untuk menyimpan hasil TF-IDF negatif
9

```

Gambar 4.2. Pseudocode Inisialisasi Variabel Global

### 4.7.3. Pembuatan fungsi getNewsFromUrl

```

1 Inisialisasi:
2 - Import library axios
3 - Import library jsdom
4 - Import readlineSync
5 - Import chalk
6 - Import JSDOM dari jsdom
7
8 Fungsi getNewsFromUrl:
9 - Minta input jumlah halaman dari pengguna
10 - Lakukan iterasi untuk setiap halaman yang diinginkan
11 - Coba mendapatkan data dari URL menggunakan axios
12 - Buat objek JSDOM dari data yang diperoleh
13 - Lakukan iterasi untuk setiap elemen berita pada halaman
14 - Ambil judul berita menggunakan selector pada objek JSDOM
15 - Tambahkan judul berita ke dalam array Judul
16 - Tingkatkan nilai o dengan 12 (offset)
17
18 Output:
19 - Tampilkan pesan atau hasil yang diinginkan setelah pengumpulan data selesai
20

```

Gambar 4.3. Pseudocode Function getNewsFromUrl

### 4.7.4. Pembuatan fungsi getBerita

```

1 - Fungsi getBerita:
2   - Tampilkan baris kosong pada konsol
3   - Periksa apakah array Judul kosong
4   - Jika kosong, tampilkan pesan bahwa judul berita kosong dan beri petunjuk untuk mengimpor CSV atau memasukkan berita
5   - Lakukan iterasi untuk setiap elemen dalam array Judul
6   - Tampilkan nomor urut dan judul berita pada konsol dengan menggunakan chalk untuk pewarnaan teks
7
8 Output:
9   - Tampilkan daftar judul berita dengan nomor urut pada konsol
10

```

Gambar 4.4. Pseudocode Function getBerita

#### 4.7.5. Pembuatan fungsi importFromCSV

```

1 Fungsi importFromCSV:
2   - Menerima parameter 'file' yang berisi path file CSV yang akan diimpor
3   - Coba membaca isi file dengan menggunakan fs.readFileSync
4   - Pisahkan data menjadi baris menggunakan '\n'
5   - Lakukan iterasi untuk setiap baris, dimulai dari indeks 1 untuk menghindari header
6   - Hapus karakter kutip ganda dan spasi ekstra, simpan hasilnya sebagai judul
7   - Tambahkan judul ke dalam array Judul
8   - Tampilkan pesan sukses jika impor berhasil
9   - Tangani kesalahan dan tampilkan pesan error jika terjadi
10
11 Output:
12 - Tampilkan pesan sukses atau pesan error tergantung pada hasil impor dari file CSV
13

```

Gambar 4.5. Pseudocode Function importFromCSV

#### 4.7.6. Pembuatan fungsi preprocessTextAndSaveToCSV

```

1 Fungsi preprocessTextAndSaveToCSV:
2   - Periksa apakah array Judul tidak kosong
3   - Jika tidak kosong, lanjutkan proses
4   - Jika kosong, tampilkan pesan bahwa judul berita kosong dan beri petunjuk untuk mengimpor CSV atau memasukkan berita
5   - Baca file custom stopwords dan kata dasar, dan proses data berdasarkan stopwords dan kata dasar
6   - Lakukan iterasi untuk setiap judul dalam array Judul
7   - Tokenisasi judul menjadi array kata-kata
8   - Lakukan preprocessing pada setiap token dengan menghapus karakter non-alphabet, menghapus stopwords, dan mengambil kata dasar
9   - Gabungkan token yang sudah diproses menjadi string dan tampilkan pada konsol
10  - Simpan hasil preprocessing ke dalam array preprocessedData
11  - Filter token yang tidak kosong dari preprocessedData
12  - Simpan hasil preprocessing ke dalam file CSV menggunakan fungsi saveToCSV
13  - Tampilkan pesan sukses bahwa data yang sudah diproses berhasil disimpan ke preprocessed_berita.csv
14
15 Output:
16 - Tampilkan hasil preprocessing pada konsol
17 - Tampilkan pesan sukses atau pesan error tergantung pada hasil pemrosesan dan penyimpanan ke dalam file CSV
18

```

Gambar 4.6. Pseudocode Function preprocessTextAndSaveToCSV

#### 4.7.7. Pembuatan fungsi calculateTFIDFForCombinedData

```

1 Fungsi calculateTFIDFForCombinedData:
2   - Inisialisasi variabel yang dibutuhkan seperti jumlah dokumen (documentCount), term frequency (termFrequency),
3     document frequency (documentFrequency), inverse document frequency (inverseDocumentFrequency), dan unik kata (uniqueTerms)
4   - Lakukan iterasi untuk setiap data dalam combinedData
5   - Tokenisasi teks dari setiap data dan pisahkan menjadi token
6   - Lakukan iterasi untuk setiap token
7   - Bersihkan token dari karakter tambahan dan spasi
8   - Hitung term frequency (TF) untuk setiap dokumen
9   - Tambahkan kata unik ke dalam set uniqueTerms
10  - Hitung term frequency untuk setiap term, document frequency, dan tambahkan ke termFrequency dan documentFrequency
11  - Hitung inverse document frequency (IDF) untuk setiap term
12  - Inisialisasi array tfidfData untuk menyimpan hasil perhitungan TF-IDF
13  - Lakukan iterasi untuk setiap term dalam uniqueTerms
14  - Lakukan iterasi untuk setiap dokumen
15  - Hitung TF, DF, IDF, dan TF-IDF
16  - Tambahkan hasil perhitungan ke dalam array tfidfData
17  - Kembalikan array tfidfData sebagai hasil perhitungan TF-IDF
18
19 Output:
20 - Array tfidfData yang berisi hasil perhitungan TF-IDF untuk setiap term dan dokumen dalam combinedData
21

```

Gambar 4.7. Pseudocode Function calculateTFIDFForCombinedData



#### 4.7.8. Pembuatan fungsi readTrainingAndTestDataFromCSV

```

1 Fungsi readTrainingAndTestDataFromCSV:
2   - Inisialisasi variabel seperti preprocessedData, dfSentimenPositive, dan dfSentimenNegative
3   - Baca dan proses training data
4     - Baca dan tokenisasi setiap baris dari file training
5     - Simpan data training yang sudah diproses ke dalam preprocessedData dan trainingData
6     - Hitung dfSentimenPositive dan dfSentimenNegative untuk setiap term
7   - Baca dan proses test data
8     - Baca dan tokenisasi setiap baris dari file test
9     - Simpan data test yang sudah diproses ke dalam preprocessedData dan testData
10  - Gabungkan training dan test data ke dalam combinedData
11  - Hitung TF-IDF untuk combinedData menggunakan fungsi calculateTFIDFForCombinedData
12  - Lampirkan dfSentimenPos dan dfSentimenNeg ke setiap item di combinedData
13  - Buat output CSV untuk data training, data test, dan data combined
14  - Hitung total TF-IDF Positive dan Negative untuk data training
15  - Simpan hasil perhitungan TF-IDF ke dalam file CSV
16  - Tampilkan pesan sukses atau pesan error jika terjadi kesalahan selama proses
17
18 Output:
19   - File CSV berisi hasil perhitungan TF-IDF untuk data training, data test, dan data combined
20   - Pesan sukses atau pesan error pada konsol tergantung pada hasil eksekusi fungsi
21

```

Gambar 4.8. Pseudocode Function readTrainingAndTestDataFromCSV

#### 4.7.9. Pembuatan fungsi calculateNaiveBayesProbabilities

```

1 Fungsi calculateNaiveBayesProbabilities:
2   - Inisialisasi variabel csvData dan resultObject
3   - Inisialisasi totalV sebagai jumlah total term unik
4   - Lakukan iterasi untuk setiap data dalam testData
5     - Tokenisasi teks dari setiap data dan ubah menjadi array
6     - Inisialisasi uniqueTermsSet untuk menyimpan term unik
7   - Lakukan iterasi untuk setiap item dalam combinedData
8     - Lakukan iterasi untuk setiap TF-IDF item dalam item
9       - Tambahkan setiap term ke uniqueTermsSet
10  - Hitung total jumlah term unik
11  - Lakukan iterasi untuk setiap data dalam testData
12    - Inisialisasi probabilitas positif dan negatif untuk setiap dokumen
13  - Lakukan iterasi untuk setiap TF-IDF item dalam data
14    - Hitung probabilitas positif dan negatif untuk setiap term
15    - Hitung total probabilitas positif dan negatif untuk setiap dokumen
16    - Tampilkan log probabilitas untuk setiap term
17    - Tampilkan total probabilitas positif dan negatif untuk setiap dokumen
18  - Tentukan sentimen dokumen berdasarkan perbandingan probabilitas
19  - Tambahkan hasil analisis ke resultObject
20  - Tampilkan hasil analisis untuk setiap dokumen pada konsol
21  - Gabungkan hasil analisis ke dalam string CSV
22  - Simpan hasil analisis ke dalam file CSV
23  - Tampilkan pesan sukses atau pesan error jika terjadi kesalahan selama proses
24 Output:
25   - File CSV berisi hasil analisis sentimen untuk setiap dokumen dalam testData
26   - Pesan sukses atau pesan error pada konsol tergantung pada hasil eksekusi fungsi
27

```

Gambar 4.9. Pseudocode Function calculateNaiveBayesProbabilities

#### 4.7.10. Pembuatan fungsi readSentimenResult

```

1 Fungsi readSentimenResult:
2   - Menerima parameter 'filename', yang berisi path dari file CSV yang akan dibaca
3   - Coba membaca isi file CSV
4   - Jika berhasil, proses data
5     - Pisahkan baris-baris menjadi array
6     - Inisialisasi array dataResult untuk menyimpan hasil pembacaan
7   - Lakukan iterasi untuk setiap baris dalam array
8     - Pisahkan setiap baris menjadi array data dengan menghapus karakter tambahan dan spasi
9     - Jika panjang array lebih dari 1, proses data
10      - Ambil judul dari array dengan menghapus karakter kutip ganda dan spasi
11      - Ambil sentimen prediksi dan sentimen kebenaran dari array
12      - Tambahkan objek baru ke dalam array dataResult
13   - Kembalikan array dataResult sebagai hasil pembacaan
14   - Jika terjadi kesalahan selama proses membaca, tangkap error dan tampilkan pesan error
15   - Kembalikan array kosong sebagai hasil pembacaan
16

```

Gambar 4.10. Pseudocode Function readSentimenResult

#### 4.7.11. Pembuatan fungsi calculateMetrics

```

1 Fungsi calculateMetrics:
2   - Inisialisasi variabel truePositives, trueNegatives, falsePositives, falseNegatives
3   - Inisialisasi sentimenAnalisis sebagai nama file 'HasilAnalisisSentimen.csv'
4   - Cek apakah file 'HasilAnalisisSentimen.csv' tersedia
5   - Jika file tersedia, lakukan iterasi untuk setiap item dalam dataResult
6     - Evaluasi sentimen prediksi dan sentimen kebenaran untuk setiap item
7     - Perbarui variabel truePositives, trueNegatives, falsePositives, falseNegatives
8   - Hitung akurasi, presisi, dan recall
9   - Tampilkan confusion matrix dan hasil evaluasi pada konsol
10  - Tampilkan pesan sukses pada konsol
11  - Tampilkan menu (jika ada)
12  - Jika file tidak ditemukan, tampilkan pesan error pada konsol
13  - Tampilkan menu (jika ada)
14

```

Gambar 4.11. Pseudocode Function calculateMetrics

#### 4.7.12. Pembuatan fungsi menu

```

1 Fungsi menu:
2   - Tampilkan menu utama pada konsol
3   - Baca input dari pengguna untuk memilih menu
4   - Gunakan switch case untuk memproses pilihan menu
5   - Jika pilihan menu adalah "1", panggil fungsi getNewsFromURL dan kembali ke menu
6   - Jika pilihan menu adalah "2", panggil fungsi getBerita dan kembali ke menu
7   - Jika pilihan menu adalah "3", simpan data berita ke dalam file CSV jika tersedia, atau tampilkan pesan error jika tidak ada data
8   - Jika pilihan menu adalah "4", baca input file dari pengguna dan panggil fungsi importFromCSV, kembali ke menu
9   - Jika pilihan menu adalah "5", panggil fungsi preprocessTextAndSaveToCSV dan kembali ke menu
10  - Jika pilihan menu adalah "6", panggil fungsi readTrainingAndTestDataFromCSV, calculateNaiveBayesProbabilities, dan kembali ke menu
11  - Jika pilihan menu adalah "7", panggil fungsi calculateMetrics dengan hasil pembacaan sentimen dari "HasilAnalisisSentimen.csv"
12  - Jika pilihan menu tidak valid, kembali ke menu
13
14 Output:
15   - Tampilkan menu utama dan hasil eksekusi menu yang dipilih oleh pengguna
16

```

Gambar 4.12. Pseudocode Function menu

## 4.8. Hasil Program

Setelah dilakukan pengembangan program klasifikasi sentimen, berikut merupakan hasil tangkapan layar dari program klasifikasi sentimen saat dijalankan:

### 4.8.1. Tampilan Menu Program

Pada tampilan menu, terdapat beberapa pilihan fungsi untuk memudahkan saat penelitian berlangsung

```

Microsoft Windows [Version 10.0.22631.2861]
(c) Microsoft Corporation. All rights reserved.

C:\Users\User>cd/skripsi
C:\skripsi>node skripsi

SKRIPSI

===== MENU =====
[1] Input news
[2] Get news
[3] Save CSV
[4] Import CSV
[5] Preprocessing
[6] Calculate TF-IDF - Train Dataset - NBC Classification
[7] Confussion Matrix
=====
[+] Menu :

```

Gambar 4.13. Tampilan Menu Program

### 4.8.2. Hasil Scrape Judul Berita

Pada bagian program scrape judul berita, dengan *input* (1) untuk melakukan intruksi total scraping data, kemudian dilakukan *input* (2) untuk menampilkan hasil scrape dan *input* (3) untuk menyimpan kedalam csv

```

1996. Link Live Streaming PSIS Vs Madura United Jam 15.00 WIB: Mahesa
Jenar Ingin Menang di Akhir Tahun
1997. Antusias Peserta Semarang 10K saat Pengambilan Race Pack di Ged
ung Oudetrap
1998. Mbak Ita Dorong Anak Tanam Jiwa Wirausaha Sejak Dini
1999. Pelaku Curanmor di Balaikota Semarang Tertangkap
2000. Perempuan Pembuang Bayi di Bawah Jembatan Diamankan Polisi
2001. Pemilu yang Jujur dan Adil | Zona Inspirasi
2002. Hoaks Politik Jelang Pemilu 2024 Naik 100 Persen
2003. Hobi Unik Koleksi Diecast 87
2004. Libur Nataru, Konsumsi BBM diperkirakan Naik 7%
2005. BPN Jawa Tengah Targetkan Sisa PTSL Selesai Tahun Depan
2006. Waspada Bencana, Tim SAR Gelar Apel Siaga

===== MENU =====
[1] Input news
[2] Get news
[3] Save CSV
[4] Import CSV
[5] Preprocessing
[6] Calculate TF-IDF - Train Dataset - NBC Classification
[7] Confussion Matrix
=====
[+] Menu :

```

Gambar 4.14. Tampilan Program Scrape Judul Berita

#### 4.8.3. Hasil Program Import Csv

Pada program import csv, dengan input (4) dan nama berkas csv yang mana isi csv ini berupa *dataset* judul berita

```
===== MENU =====
[1] Input news
[2] Get news
[3] Save CSV
[4] Import CSV
[5] Preprocessing
[6] Calculate TF-IDF - Train Dataset - NBC Classification
[7] Confussion Matrix
=====

[+] Menu : 4
[+] Input file : berita.csv
Data berhasil diimpor dari file CSV.
```

Gambar 4.15. Tampilan Program *Import* Judul Berita

#### 4.8.4. Hasil Preprocessing Judul Berita

Pada bagian program *preprocessing* dengan *input* (5) terdapat fungsi untuk melakukan beberapa proses yakni, *cleansing*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil preprocessing akan disimpan kedalam file csv.

```
Command Prompt - node sks X + -
'temu,salur,buang,jenazah,ambang,labuh',
'bacok,guru,bantu,keluarga,jual',
'sosok,mayat,pria,misterius,labuh,tanjung,emas,temu',
'sosok,mayat,pria,tanpa,identitas,apung,labuh,tanjung,emas',
'siap,aman,gelar,simulasi',
'bantu,perintah,belum,normal,harga',
'harga,borong',
'pamer,sehat,suguh,aneka,layan,fasilitas,sehat',
'nal,senam,tera,olah,raga,usia',
'tinjau,harga,harga,tinggi',
... 1786 more items
]
Data yang sudah diproses berhasil disimpan ke preprocessed_berita.csv

===== MENU =====
[1] Input news
[2] Get news
[3] Save CSV
[4] Import CSV
[5] Preprocessing
[6] Calculate TF-IDF - Train Dataset - NBC Classification
[7] Confussion Matrix
=====

[+] Menu :
```

Gambar 4.16. Tampilan Program *Preprocessing* Judul Berita

#### 4.8.5. Hasil Perhitungan TF-IDF dan Klasifikasi NBC

Setelah dilakukan *preprocessing* didapatkan term di setiap dokumen beserta frekuensi kemunculan data kemudian dihitung TF-IDF, dan akan dilakukan perhitungan probabilitas dengan rumus Naïve Bayes Classifier, hasil klasifikasi akan disimpan kedalam file csv

```

Command Prompt - node sks
1886. Dokumen ke-1886 (titik,banjir,bagi,mulai,surut,masuk,pertama)
P(bagi|Positif)=(25+1)/(9310.32691+1999)=0.00279
P(bagi|Negatif)=(6+1)/(8772.84415+1999)=0.00080
P(mulai|Positif)=(14+1)/(9310.32691+1999)=0.00161
P(mulai|Negatif)=(11+1)/(8772.84415+1999)=0.00137
P(masuk|Positif)=(6+1)/(9310.32691+1999)=0.00075
P(masuk|Negatif)=(3+1)/(8772.84415+1999)=0.00046
P(pertama|Positif)=(6+1)/(9310.32691+1999)=0.00075
P(pertama|Negatif)=(0+1)/(8772.84415+1999)=0.00011
P(banjir|Positif)=(9+1)/(9310.32691+1999)=0.00107
P(banjir|Negatif)=(13+1)/(8772.84415+1999)=0.00160
P(titik|Positif)=(0+1)/(9310.32691+1999)=0.00011
P(titik|Negatif)=(1+1)/(8772.84415+1999)=0.00023
P(surut|Positif)=(0+1)/(9310.32691+1999)=0.00011
P(surut|Negatif)=(0+1)/(8772.84415+1999)=0.00011
Total P( Positif ) = 0.00720
Total P( Negatif ) = 0.00467
Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-1886 bersentimen positif
Results saved to Klasifikasi HasilAnalisisSentimen.csv

===== MENU =====
[1] Input news
[2] Get news
[3] Save CSV
[4] Import CSV
[5] Preprocessing
[6] Calculate TF-IDF - Train Dataset - NBC Classification

```

Gambar 4.17. Tampilan Program Perhitungan TF-IDF &amp; Klasifikasi NBC

#### 4.8.6. Hasil Perhitungan Confussion Matrix

Dalam program ini, dengan *input* (7) dilakukan perhitungan performa metode Naïve Bayes Classifier menggunakan confusion matrix. Selanjutnya akan menampilkan hasil performa Naïve Bayes Classifier

```

Command Prompt - node sks
-----
| Prediksi Sentimen | Sentimen Asli | |
|---|---|---|
| POSITIF | 825 | 57 |
| NEGATIF | 112 | 892 |
|-----|-----|

Confusion Matrix:
- True Positives: 825
- True Negatives: 892
- False Positives: 57
- False Negatives: 112
Result:
- Akurasi: 91.04 %
- Presisi Kelas Negatif: 0.89
- Recall Kelas Negatif: 0.94
- Presisi Kelas Positif: 0.94
- Recall Kelas Positif: 0.88

===== MENU =====
[1] Input news
[2] Get news
[3] Save CSV
[4] Import CSV
[5] Preprocessing
[6] Calculate TF-IDF - Train Dataset - NBC Classification
[7] Confusion Matrix
=====

```

Gambar 4.18. Tampilan Program Perhitungan Confussion Matrix

## BAB V PENUTUP

### 5.1. Kesimpulan

Setelah mendapatkan hasil penelitian tentang klasifikasi sentimen pada judul berita Kompas.Tv Kota Semarang, maka didapatkan kesimpulan sebagai berikut:

1. Data yang didapat harus melalui proses *preprocessing* sebelum pengolahan data menggunakan algoritma Naïve Bayes Classifier. Proses ini memiliki tujuan menormalisasikan data sehingga proses pengolahan data mendapatkan hasil akurasi yang tinggi.
2. Pada proses *stopword removal*, banyak kata yang tidak memiliki makna atau sentimen, sehingga berpengaruh terhadap hasil klasifikasi menggunakan metode Naïve Bayes Classifier.
3. Kata “laku” merupakan kata yang paling banyak keluar di frekuensi dokumen negatif dengan total kemunculan kata 79 dan tujuh di frekuensi dokumen positif.
4. Pada hasil program klasifikasi judul berita pada Kompas Tv Kota Semarang yang telah berhasil diklasifikasi oleh program mayoritas bersentimen negatif.
5. Hasil penelitian menggunakan algoritma Naïve Bayes Classifier pada judul berita pada Kompas Tv Kota Semarang memiliki performa akurasi sebesar 91.04%, presisi kelas negatif sebesar 89% dan *recall* kelas negatif sebesar 94% sedangkan presisi kelas positif sebesar 94% dan *recall* kelas positif sebesar 88%.

### 5.2. Saran

Penelitian ini membuka peluang bagi peneliti masa depan untuk melakukan pengembangan lebih lanjut dalam bidang klasifikasi sentimen pada judul berita online menggunakan metode Naive Bayes Classifier. Berikut merupakan saran dari peneliti:

1. Bagi peneliti yang akan datang dilakukan evaluasi mendalam terhadap kamus *stopwords* yang digunakan pada proses *stopword removal*. Proses

evaluasi ini bertujuan untuk memastikan bahwa kata-kata yang tidak memiliki makna atau sentimen yang signifikan dapat terfilter dengan baik.

2. Bagi peneliti yang akan datang dengan topik pengklasifikasian sentimen judul berita dapat menggunakan algoritma yang lain serta memperkaya data latih.

