

**SKRIPSI**

**PENERAPAN METODE *NAIVE BAYES CLASSIFIER* UNTUK**

**KLASIFIKASI SENTIMEN PADA JUDUL BERITA**

**KOMPAS.TV KOTA SEMARANG**

Diajukan untuk memenuhi memenuhi salah satu syarat mendapatkan  
gelar Sarjana Teknik Informatika



Disusun Oleh:

Nama : Alrico Rizki Wibowo

NIM : A11.2020.12703

Program Studi : Teknik Informatika

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS DIAN NUSWANTORO**

**2024**

## UCAPAN TERIMA KASIH

Segala puji bagi Allah SWT yang memberikan taufik dan hidayah-Nya. Terima kasih kepada semua yang telah membantu dan membimbing dalam penulisan laporan tugas akhir yang berjudul “PENERAPAN METODE NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI SENTIMEN PADA JUDUL BERITA KOMPAS.TV KOTA SEMARANG” dapat terselesaikan. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Edi Noersasongko, M.Kom., selaku Bapak Rektor Universitas Dian Nuswantoro.
2. Bapak Dr. Abdul Syukur selaku Bapak Dekan Universitas Dian Nuswantoro.
3. Bapak Sri Winarno, Ph.D selaku Kaprogdi TI-SI.
4. Ibu Yani Parti Astuti, S.Si, M.Kom selaku pembimbing tugas akhir.
5. Kedua orang tua penulis, yang selalu memberikan semangat dan dukungan.
6. Sri Mulyati, S.S. dan Dewi Fitri Titasari, S.S., yang membantu dalam pembuatan label klasifikasi *dataset*.
7. Silvianita Ika Aprillia Vridayanti, S.Kep yang memberikan saran dan nasehat dalam pembuatan tugas akhir.
8. Seluruh sahabat dan teman-teman penulis angkatan 2020 yang tidak bisa disebutkan satu-persatu.

Pada akhirnya, diharapkan laporan tugas akhir ini dapat memberikan manfaat yang signifikan. Penulis menyadari bahwa dalam penyusunannya masih terdapat kekurangan yang perlu diperbaiki untuk mencapai tingkat kesempurnaan yang diinginkan.

Semarang, 15 Januari 2024



Alrico Rizki Wibowo

A11.2020.12703

## **ABSTRAK**

Berita memiliki peran utama sebagai sumber informasi untuk menyampaikan laporan mengenai pendapat, peristiwa, dan temuan terbaru dalam berbagai aspek kehidupan. Judul berita, sebagai komponen penting, mampu menjadi penentu dari isi berita. Sentimen yang terkandung dalam judul berita dapat diklasifikasikan dengan menggunakan sentimen analis, seperti yang terjadi di platform media online Kompas.TV. Judul berita diambil dengan menggunakan program otomatis yang memanfaatkan body HTML dengan bantuan NodeJs sebagai teknologi untuk pembuatan program. Penelitian ini difokuskan pada penerapan metode Naïve Bayes Classifier untuk mengklasifikasikan sentimen pada judul berita Kompas.TV di Kota Semarang. Hasil penelitian menunjukkan tingkat akurasi mencapai 91.04%, dengan rasio data latih dan data uji sebesar 90:10. Kesimpulan dari penelitian ini adalah bahwa metode Naïve Bayes Classifier efektif dalam mengidentifikasi judul berita dengan sentimen negatif pada Kompas.TV, dengan presisi sebesar 89% dan recall sebesar 94%. Temuan ini memberikan kontribusi positif pada pemahaman analisis sentimen pada judul berita di media online, terutama dalam konteks berita Kompas.TV di Kota Semarang.

Kata Kunci : Kompas.TV, Analisis Sentimen, Naïve Bayes Classifier

## DAFTAR ISI

UCAPAN TERIMA KASIH.....	II
ABSTRAK.....	III
DAFTAR ISI.....	IV
DAFTAR TABEL.....	VI
DAFTAR GAMBAR .....	VII
BAB I.....	1
PENDAHULUAN .....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	4
BAB II.....	5
TINJAUAN PUSTAKA .....	5
2.1. Tinjauan Studi .....	5
2.2. Tinjauan Pustaka .....	7
2.3. Kerangka Pemikiran.....	13
BAB III .....	15
METODOLOGI PENELITIAN.....	15
3.1. Sumber Data.....	15
3.2. Instrumen Penelitian.....	15
3.3. Metode Yang Diusulkan .....	16
BAB IV .....	22
HASIL PENELITIAN DAN PEMBAHASAN .....	22
4.1. Pengumpulan Data .....	22
4.2. Tahap Preprocessing Data.....	23
4.3. Pembobotan Kata .....	32
4.4. Klasifikasi Metode .....	37

4.5. Pengujian Metode.....	40
4.6. Hasil Evaluasi.....	41
4.7. Pseudocode Program.....	41
4.8. Hasil Program .....	49
BAB V PENUTUP.....	54
5.1. Kesimpulan .....	54
5.2. Saran.....	54
DAFTAR PUSTAKA .....	72

## DAFTAR TABEL

Tabel 2.1 Penelitian Terkaiit .....	6
Tabel 3.1. Tabel Data Set .....	18
Tabel 3.2. Cleansing .....	18
Tabel 3.3. <i>Case Folding</i> .....	19
Tabel 3.4. Tokenizing .....	19
Tabel 3.5 Stopword Removal.....	20
Tabel 3.6. <i>Stemming</i> .....	20
Tabel 3.8 Confussion Matrix .....	21
Tabel 4.1 Contoh Dataset .....	22
Tabel 4.2 Contoh Pelabelan Dataset .....	23
Tabel 4.3.1 Hasil Sampel Dataset <i>Cleansing</i> .....	24
Tabel 4.3.2 Hasil Sampel Data <i>Set Case Folding</i> .....	26
Tabel 4.3.3 Hasil Sampel Dataset <i>Tokenizing</i> .....	27
Tabel 4.3.4 Hasil Sampel Dataset <i>Stopword Removal</i> .....	29
Tabel 4.3.5 Hasil Sampel Dataset <i>Stemming</i> .....	30
Tabel 4.4.1 Hasil TF-IDF Sampel Dataset <i>Training</i> .....	32
Tabel 4.4.1 Hasil TF-IDF Sampel Dataset <i>Testing</i> .....	34
Tabel 4.6.1 Perbandingan Hasil Performa NBC Rasio Dataset .....	41

## DAFTAR GAMBAR

Gambar 2.1. Kerangka Pemikiran .....	14
Gambar 3.1. Alur Penelitian.....	17
Gambar 4.1. Pseudocode Inisialisasi Import Library .....	42
Gambar 4.2. Pseudocode Inisialisasi Variabel Global .....	42
Gambar 4.3. Pseudocode Function getNewsFromUrl.....	43
Gambar 4.4. Pseudocode Function getBerita .....	43
Gambar 4.5. Pseudocode Function importFromCSV .....	44
Gambar 4.6. Pseudocode Function preprocessTextAndSaveToCSV .....	45
Gambar 4.7. Pseudocode Function calculateTFIDF .....	45
Gambar 4.8. Pseudocode Function readTrainingAndTestDataFromCSV .....	46
Gambar 4.9. Pseudocode Function calculateNaiveBayesProbabilities.....	47
Gambar 4.10. Pseudocode Function readSentimenResult .....	48
Gambar 4.11. Pseudocode Function calculateMetrics .....	48
Gambar 4.12. Pseudocode Function menu.....	49
Gambar 4.13. Tampilan Menu Program .....	50
Gambar 4.14. Tampilan Program Scrape Judul Berita .....	50
Gambar 4.15. Tampilan Program <i>Import</i> Judul Berita .....	51
Gambar 4.16. Tampilan Program <i>Preprocessing</i> Judul Berita .....	51
Gambar 4.17. Tampilan Program Perhitungan TF-IDF & Klasifikasi NBC .....	52
Gambar 4.18. Tampilan Program Perhitungan Confussion Matrix .....	53

# **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 sentimen netral sebanyak 147 data, sentimen positif sebanyak 114 data, dan sentimen 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 yang sama.

### **1.2. Rumusan Masalah**

Dengan merujuk terhadap 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 Januari 2023 sampai November 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 *Naïve 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 calon presiden dan wakil 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". Didapatkan tingkat akurasi sebesar 80% pada 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. Memiliki 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 Pemilu 2019	<i>Naïve Bayes Classifier</i>	Didapatkan 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	Analisis klasifikasi pada judul berita online	<i>Naïve Bayes Classifier</i>	Dalam penelitian ini, diperoleh 400 data berita yang diklasifikasikan ke dalam empat kelas, yaitu banua, bisnis, hukum, dan kriminal.

	Al Awalaien, Ghifari, 2021	Radar Banjarmasin.		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 COVID-19 di media daring DetikHealth pada tahun 2021	<i>Naïve Bayes Classifier</i>	Dari hasil pengumpulan data 399 judul berita, dikelompokkan ke dalam tiga kategori sentimen, yakni sentimen netral dengan 147 data, sentimen positif sebanyak 114 data, dan sentimen negatif sejumlah 138 data. Memiliki 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 Perdana, Rizal Setya, 2023	Klasifikasi judul berita <i>online</i> untuk membuat kategori berita	<i>Support Vector Machine</i>	Penelitian ini menghasilkan tingkat akurasi sebesar 93,06%, dengan presisi mencapai 92,11%, recall 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* adalah 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* adalah suatu 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:

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

Keterangan :

$P(C)$  : Probabilitas dari kelas C

$P(W_i, C)$  : Jumlah *terms* yang ada di dokumen data latih pada kelas C

$\text{Count}(C)$  : Jumlah semua *terms* yang ada di dokumen data latih pada kelas C

$|V|$  : Jumlah total *terms* (*term* yang sama dihitung 1)

### 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.

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

Keterangan:

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

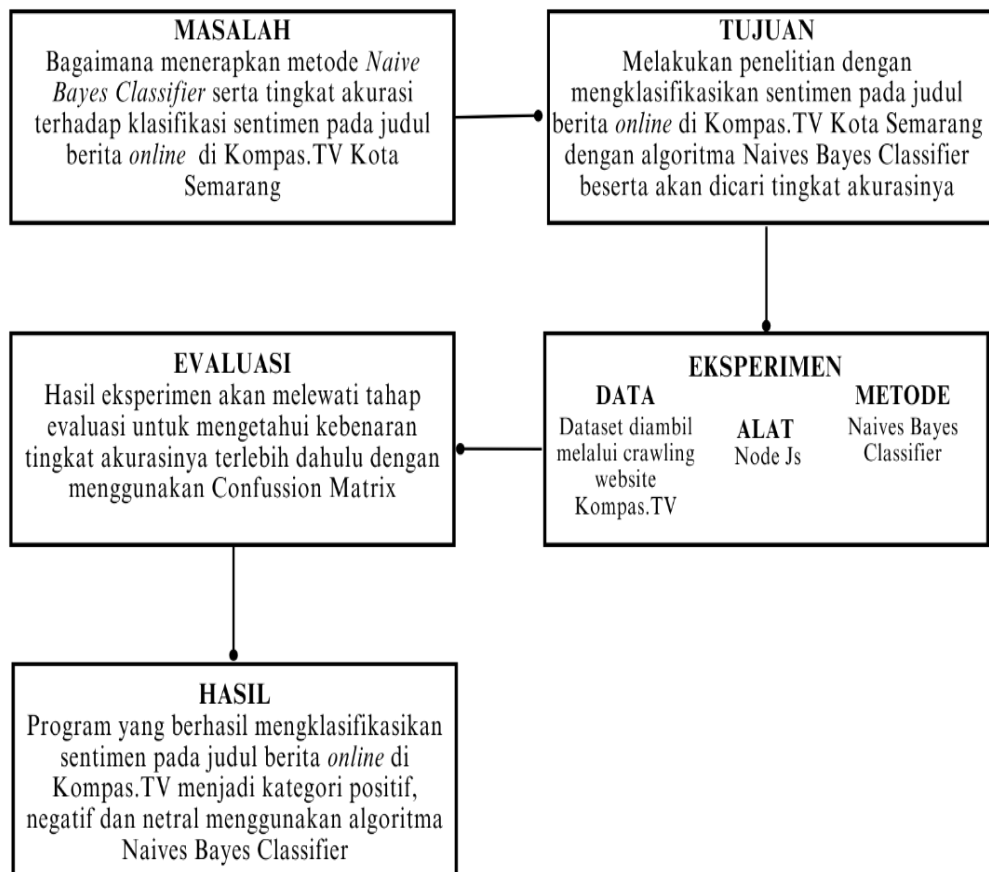
### 2.2.9. Node JS

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, SastrawiJs dan Axios.

### 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.

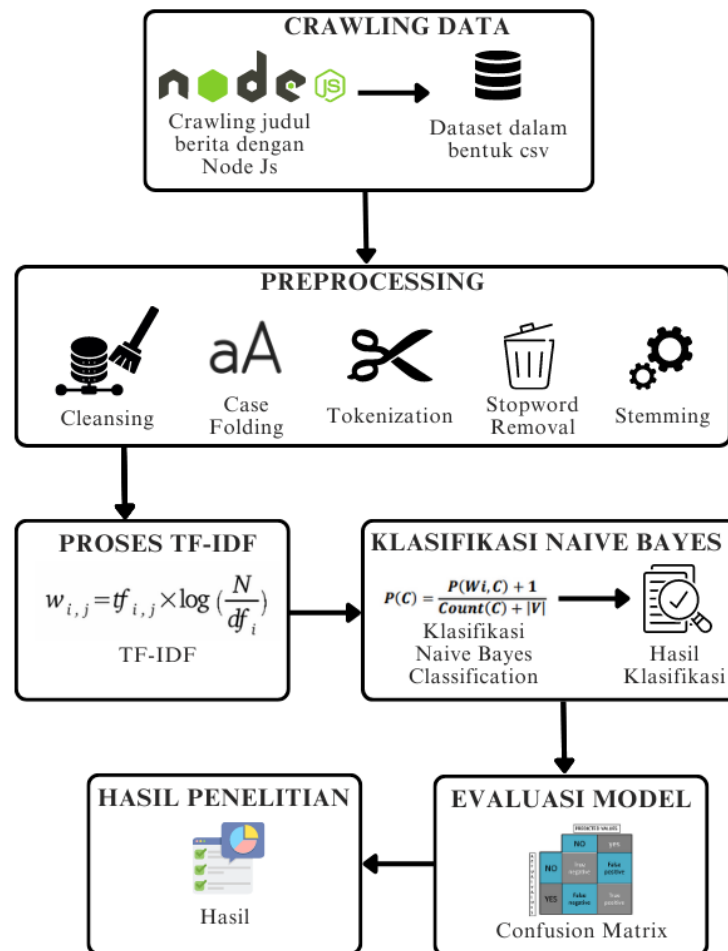
### 3.3. Metode Yang Diusulkan

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.

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. Pelabelan Dataset

Pelabelan dataset akan dilakukan setelah data terkumpul, dimana pelabelan ini akan dilabeli secara manual oleh penganalisis yang telah menempuh gelar sarjana sastra. Pelabelan ini berfungsi untuk mengukur hasil prediksi dari program yang akan dibuat menggunakan metode pengklasifikasian Naïve Bayes Classifier.



### 3.3.3. 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

### 3.3.4. 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.

### 3.3.5. 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*.

### 3.3.6. 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

Penelitian ini diawali dengan tahap pengumpulan data berupa teks dari judul berita Kompas TV yang berkaitan dengan wilayah Kota Semarang. Proses pengambilan data dilakukan menggunakan teknik pengambilan data (crawling) yang dikembangkan dengan menggunakan Node JS. Dalam implementasinya, library seperti axios dan jsdom digunakan untuk melakukan *parsing* HTML dan mengakses informasi yang dibutuhkan. Total dataset yang berhasil dikumpulkan sebanyak 1886 judul berita. Untuk keperluan analisis, dataset tersebut kemudian dibagi menjadi dua bagian, yakni data training sebesar 90% dan data testing sebesar 10%, guna memastikan representasi yang seimbang dalam tahap pembelajaran dan pengujian model..

Tabel 4.1 Contoh Dataset

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

## 4.2 Pelabelan Dataset

Setelah berhasil mengumpulkan data, dilakukan analisis sentimen serta pelabelan data secara manual oleh penganalisis yang telah memperoleh gelar sarjana sastra Bahasa Indonesia. Langkah ini menunjukkan upaya untuk memberikan label pada setiap dataset berdasarkan sentimen yang terkandung di dalamnya. Proses pelabelan dilakukan dengan memanfaatkan pemahaman mendalam terhadap struktur bahasa dan konteks budaya Indonesia yang dimiliki oleh penganalisis yang berlatar belakang gelar sarjana sastra Bahasa Indonesia. Dengan demikian, setiap entri dalam dataset diberikan label sentimen yang mencerminkan evaluasi atau perasaan yang terkandung dalam teks tersebut. Berikut adalah contoh dataset yang telah dilakukan pelabelan :

Tabel 4.2 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

## 4.3. 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*, *stopword removal*, dan *stemming*. Berikut adalah tabel hasil dari proses-proses yang dilakukan :

#### 4.3.1. Cleansing

Tahap cleansing memiliki fokus utama pada penghapusan atau penanganan simbol-simbol yang tidak diperlukan atau mungkin mengganggu analisis data. Proses ini mencakup identifikasi dan penghapusan karakter khusus, tanda baca, atau simbol lainnya yang tidak memberikan kontribusi signifikan terhadap analisis teks. Langkah-langkah dalam cleansing dapat mencakup filterisasi karakter non-alfanumerik, penghilangan tanda baca, dan pemrosesan khusus terhadap simbol-simbol tertentu yang dianggap tidak relevan. Hasil dari tahap cleansing ini dapat diamati dalam Tabel 4.3.1

Tabel 4.3.1 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 Dropped 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

#### 4.3.2. Case Folding

Case folding menjadi langkah berikutnya setelah cleansing, di mana setiap entri dalam tabel mencerminkan kata-kata atau teks yang telah diubah menjadi huruf kecil. Dengan menerapkan *case folding*, konsistensi dalam penulisan kata-kata dijaga, sehingga kata-kata dengan varian huruf besar dan kecil dianggap setara



Tabel 4.3.2 Hasil Sampel Data *Set 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 dropping 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 dpmpstsp 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

#### 4.3.3. Tokenizing

Proses tokenizing merupakan tahap selanjutnya setelah case folding, di mana teks dibagi menjadi unit-unit yang lebih kecil yang disebut sebagai token. Setiap entri dalam tabel merepresentasikan token-token yang dihasilkan dari pemisahan teks. Proses ini memecah teks menjadi bagian-bagian yang lebih kecil, membantu dalam memahami struktur kalimat dan kata-kata yang ada dalam dataset. Dengan menerapkan tokenizing, dataset menjadi lebih terstruktur, dan analisis teks dapat dilakukan lebih efisien karena fokus pada unit-unit terkecil. Ini membantu menyederhanakan kompleksitas teks, memungkinkan pemahaman yang lebih baik terhadap makna kata-kata dan kalimat dalam dataset. Proses tokenizing menjadi langkah penting dalam persiapan dataset untuk analisis lebih lanjut, memberikan dasar yang kuat untuk langkah-langkah preprocessing data berikutnya.

Tabel 4.3.3 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', 'dpmpptsp', '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'

#### 4.3.4. Stopword Removal

Langkah stopwords removal merupakan elemen kunci dari tahap preprocessing data setelah tokenizing, di mana kata-kata umum yang tidak memberikan nilai signifikan terhadap analisis dihilangkan dari dataset. Setiap entri dalam Tabel 4.3.4 mencerminkan dataset setelah proses penghapusan stopwords. Dengan implementasi stopwords removal, dataset menjadi lebih terfokus pada kata-kata kunci yang membawa informasi yang lebih relevan dalam analisis, signifikan meningkatkan ketepatan dan keakuratan analisis teks. Tabel 4.3.4 memberikan visualisasi yang jelas mengenai perubahan dataset setelah langkah stopwords removal, menunjukkan bagaimana proses ini berkontribusi dalam menyederhanakan dataset untuk analisis lebih lanjut, serta memastikan bahwa dataset yang digunakan lebih terfokus pada informasi yang substansial. Proses stopwords removal menjadi langkah kritis dalam mempersiapkan dataset untuk analisis teks yang lebih akurat dan informatif.

Tabel 4.3.4 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

#### 4.3.5. Stemming

Tabel 4.3.5 menggambarkan hasil dari proses *stemming* dengan kamus Bahasa Indonesia pada sampel dataset setelah tahap *stopword removal*. Langkah stemming ini, yang mereduksi kata-kata menjadi bentuk dasarnya, memanfaatkan kamus bahasa Indonesia untuk memastikan bentuk kata dasar asli. Setiap entri dalam tabel mencerminkan dataset setelah implementasi *stemming*, yang bertujuan untuk menyederhanakan variasi kata-kata dan mendapatkan akar kata.

Tabel 4.3.5 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

#### 4.4. Pembobotan Kata

Tahap selanjutnya dalam proses adalah penerapan TF-IDF, di mana dataset mengalami pembobotan kata dan menampilkan total terms dokumen berdasarkan sentimennya. Dataset sampel untuk training dan testing telah menjalani proses pembobotan kata, di mana pada data training akan terlihat total term di dokumen sesuai dengan kelasnya. Berikut merupakan dataset sampel untuk *training* dan *testing* yang sudah melalui proses pembobotan kata:

Tabel 4.4.1 Hasil TF-IDF Sampel Dataset *Training*

No	Term	TF	DF	IDF	TF-IDF	Total Terms di Dokumen Positif	Total Terms di Dokumen Negatif
1	cemburu	1	1	1.47712	1.47712	0	1
2	suami	1	2	1.17609	1.17609	0	2
3	nekat	1	2	1.17609	1.17609	0	2
4	aniaya	1	2	1.17609	1.17609	0	2
5	istri	1	2	1.17609	1.17609	0	2
6	tewas	1	4	0.87506	0.87506	0	4
7	kaca	1	1	1.47712	1.47712	0	1
8	celaka	1	2	1.17609	1.17609	0	2
9	api	1	2	1.17609	1.17609	0	2
10	kendara	1	1	1.47712	1.47712	0	1
11	mogok	1	1	1.47712	1.47712	0	1
12	rel	1	1	1.47712	1.47712	0	1
13	sinau	1	1	1.47712	1.47712	0	1
14	tangki	1	2	1.17609	1.17609	0	2
15	maut	1	1	1.47712	1.47712	0	1
16	jadi	1	1	1.47712	1.47712	0	1
17	sangka	1	1	1.47712	1.47712	0	1
18	edukasi	1	1	1.47712	1.47712	1	0
19	sehat	1	2	1.17609	1.17609	2	0
20	paguyuban	1	1	1.47712	1.47712	1	0
21	gelar	1	1	1.47712	1.47712	1	0
22	kemarau	1	2	1.17609	1.17609	0	2
23	ganggu	1	1	1.47712	1.47712	0	1
24	rakit	1	1	1.47712	1.47712	0	1
25	jual	1	2	1.17609	1.17609	0	2
26	senjata	1	1	1.47712	1.47712	0	1
27	ilegal	1	3	1	1	0	3
28	harga	1	1	1.47712	1.47712	0	1

29	naik	1	3	1	1	1	2
30	tengkar	1	1	1.47712	1.47712	0	1
31	duga	1	1	1.47712	1.47712	0	1
32	keren	1	1	1.47712	1.47712	1	0
33	rawat	1	1	1.47712	1.47712	1	0
34	budaya	1	1	1.47712	1.47712	1	0
35	pamer	1	1	1.47712	1.47712	1	0
36	keris	1	1	1.47712	1.47712	1	0
37	krisis	1	1	1.47712	1.47712	0	1
38	bersih	2	2	1.17609	2.35218	1	1
39	salur	1	1	1.47712	1.47712	0	1
40	bantu	1	1	1.47712	1.47712	0	1
41	dampak	1	1	1.47712	1.47712	0	1
42	seru	1	1	1.47712	1.47712	1	0
43	estafet	1	1	1.47712	1.47712	1	0
44	tangkap	1	3	1	1	2	1
45	mata	1	1	1.47712	1.47712	1	0
46	tutup	1	1	1.47712	1.47712	1	0
47	optimis	1	1	1.47712	1.47712	1	0
48	investasi	1	2	1.17609	1.17609	2	0
49	aman	1	1	1.47712	1.47712	1	0
50	tani	1	2	1.17609	1.17609	2	0
51	basmi	1	1	1.47712	1.47712	1	0
52	hama	1	1	1.47712	1.47712	1	0
53	tikus	1	2	1.17609	1.17609	2	0
54	camat	1	1	1.47712	1.47712	1	0
55	nikmat	1	2	1.17609	1.17609	2	0
56	lobster	1	1	1.47712	1.47712	1	0
57	rawa	2	1	1.47712	2.95424	1	0
58	bakar	1	2	1.17609	1.17609	1	1
59	pening	1	1	1.47712	1.47712	1	0
60	depresi	1	1	1.47712	1.47712	0	1
61	pria	1	1	1.47712	1.47712	0	1
62	siaga	1	1	1.47712	1.47712	0	1
63	asap	1	1	1.47712	1.47712	0	1
64	tambang	1	1	1.47712	1.47712	0	1
65	marak	1	1	1.47712	1.47712	0	1
66	harus	1	1	1.47712	1.47712	0	1
67	legalitas	1	1	1.47712	1.47712	0	1
68	rawan	1	1	1.47712	1.47712	0	1
69	apar	1	1	1.47712	1.47712	0	1
70	emas	1	1	1.47712	1.47712	1	0
71	dukung	1	1	1.47712	1.47712	1	0



72	rekonstruksi	1	1	1.47712	1.47712	0	1
73	kasus	1	1	1.47712	1.47712	0	1
74	taksi	1	1	1.47712	1.47712	0	1
75	raih	1	1	1.47712	1.47712	1	0
76	menang	1	1	1.47712	1.47712	1	0
77	persik	1	1	1.47712	1.47712	1	0
78	manfaat	1	1	1.47712	1.47712	1	0
79	absen	1	1	1.47712	1.47712	1	0
80	gotong	1	1	1.47712	1.47712	1	0
81	sawah	1	2	1.17609	1.17609	2	0
82	gaet	1	1	1.47712	1.47712	1	0
83	investor	1	1	1.47712	1.47712	1	0
84	buat	1	1	1.47712	1.47712	1	0
85	insentif	1	1	1.47712	1.47712	1	0
86	luka	2	1	1.47712	2.95424	0	1
87	racik	1	1	1.47712	1.47712	0	1
88	edar	1	1	1.47712	1.47712	0	1
89	kosmetik	1	1	1.47712	1.47712	0	1
90	soto	1	1	1.47712	1.47712	1	0
91	organik	1	1	1.47712	1.47712	1	0
92	murah	1	1	1.47712	1.47712	1	0
TOTAL							
TOTAL DOKUMEN DI KELAS POSITIF						52	
TOTAL DOKUMEN DI KELAS NEGATIF						65	
TOTAL TERMS DATA LATIH						92	

Tabel 4.4.1 Hasil TF-IDF Sampel Dataset *Testing*

No	Term	TF	DF	IDF	TF-IDF
1	cemburu	1	1	1.47712	1.47712
2	suami	1	2	1.17609	1.17609
3	nekat	1	2	1.17609	1.17609
4	aniaya	1	2	1.17609	1.17609
5	istri	1	2	1.17609	1.17609
6	tewas	1	4	0.87506	0.87506
7	kaca	1	1	1.47712	1.47712
8	celaka	1	3	1	1
9	api	1	2	1.17609	1.17609
10	kendara	1	1	1.47712	1.47712
11	mogok	1	1	1.47712	1.47712
12	rel	1	1	1.47712	1.47712
13	sinau	1	1	1.47712	1.47712

14	tangki	1	2	1.17609	1.17609
15	maut	1	2	1.17609	1.17609
16	jadi	1	2	1.17609	1.17609
17	sangka	1	2	1.17609	1.17609
18	edukasi	1	2	1.17609	1.17609
19	sehat	1	2	1.17609	1.17609
20	paguyuban	1	1	1.47712	1.47712
21	gelar	1	1	1.47712	1.47712
22	kemarau	1	2	1.17609	1.17609
23	ganggu	1	1	1.47712	1.47712
24	rakit	1	1	1.47712	1.47712
25	jual	1	2	1.17609	1.17609
26	senjata	1	1	1.47712	1.47712
27	ilegal	1	3	1	1
28	harga	1	1	1.47712	1.47712
29	naik	1	3	1	1
30	tengkar	1	1	1.47712	1.47712
31	duga	1	1	1.47712	1.47712
32	keren	1	2	1.17609	1.17609
33	rawat	1	1	1.47712	1.47712
34	budaya	1	1	1.47712	1.47712
35	pamer	1	1	1.47712	1.47712
36	keris	1	1	1.47712	1.47712
37	krisis	1	1	1.47712	1.47712
38	bersih	2	2	1.17609	2.35218
39	salur	1	1	1.47712	1.47712
40	bantu	1	1	1.47712	1.47712
41	dampak	1	1	1.47712	1.47712
42	seru	1	1	1.47712	1.47712
43	estafet	1	1	1.47712	1.47712
44	tangkap	1	3	1	1
45	mata	1	1	1.47712	1.47712
46	tutup	1	1	1.47712	1.47712
47	optimis	1	1	1.47712	1.47712
48	investasi	1	2	1.17609	1.17609
49	aman	1	1	1.47712	1.47712
50	tani	1	2	1.17609	1.17609
51	basmi	1	1	1.47712	1.47712
52	hama	1	1	1.47712	1.47712
53	tikus	1	2	1.17609	1.17609
54	camat	1	1	1.47712	1.47712
55	nikmat	1	2	1.17609	1.17609
56	lobster	1	1	1.47712	1.47712

57	rawa	2	1	1.47712	2.95424
58	bakar	1	2	1.17609	1.17609
59	pening	1	1	1.47712	1.47712
60	depresi	1	1	1.47712	1.47712
61	pria	1	1	1.47712	1.47712
62	siaga	1	1	1.47712	1.47712
63	asap	1	1	1.47712	1.47712
64	tambang	1	1	1.47712	1.47712
65	marak	1	1	1.47712	1.47712
66	harus	1	1	1.47712	1.47712
67	legalitas	1	1	1.47712	1.47712
68	rawan	1	1	1.47712	1.47712
69	apar	1	1	1.47712	1.47712
70	emas	1	1	1.47712	1.47712
71	dukung	1	1	1.47712	1.47712
72	rekonstruksi	1	1	1.47712	1.47712
73	kasus	1	1	1.47712	1.47712
74	taksi	1	1	1.47712	1.47712
75	raih	1	1	1.47712	1.47712
76	menang	1	1	1.47712	1.47712
77	persik	1	1	1.47712	1.47712
78	manfaat	1	1	1.47712	1.47712
79	absen	1	1	1.47712	1.47712
80	gotong	1	1	1.47712	1.47712
81	sawah	1	2	1.17609	1.17609
82	gaet	1	1	1.47712	1.47712
83	investor	1	1	1.47712	1.47712
84	buat	1	2	1.17609	1.17609
85	insentif	1	1	1.47712	1.47712
86	luka	2	1	1.47712	2.95424
87	racik	1	1	1.47712	1.47712
88	edar	1	1	1.47712	1.47712
89	kosmetik	1	1	1.47712	1.47712
90	soto	1	1	1.47712	1.47712
91	organik	1	1	1.47712	1.47712
92	murah	1	1	1.47712	1.47712
93	animasi	1	1	1.47712	1.47712
94	cegah	1	1	1.47712	1.47712
95	banjir	1	1	1.47712	1.47712
96	bacok	1	1	1.47712	1.47712
97	guru	1	1	1.47712	1.47712
98	semester	1	1	1.47712	1.47712
TOTAL TERMS DATA UJI					98

#### 4.5. Klasifikasi Metode

Tahapan selanjutnya merupakan perhitungan probabilitas untuk mencari nilai probabilitas dari suatu *term* data uji berdasarkan data latih. Proses ini memungkinkan perhitungan probabilitas untuk menentukan seberapa mungkin suatu *term* terdapat dalam kelas tertentu. Implementasi rumus ini pada data uji memungkinkan penentuan probabilitas kelas yang paling mungkin, berdasarkan pembelajaran dari data latih. Berikut merupakan rumus mencari probabilitas Naïve Bayes Classifier dan perhitungan contoh sampel data uji:

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

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

$$P(\text{cemburu}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{cemburu}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{suami}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{suami}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{nekat}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{nekat}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{aniaya}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{aniaya}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{istri}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{istri}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{tewas}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{tewas}|\text{Negatif}) = \frac{4 + 1}{66 + 92} = 0.03165$$

$$\text{Total } P(\text{Positif}) = 0.04167$$

$$\text{Total } P(\text{Negatif}) = 0.12025$$

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

Setelah melalui seluruh perhitungan pada data uji, hasil dari perhitungan probabilitas sentimen dari setiap dokumen dikumpulkan dan disajikan dalam Tabel 4.5.1, yang merupakan hasil prediksi menggunakan metode Naïve Bayes Classifier. Tabel ini memuat informasi mengenai probabilitas sentimen yang diperoleh untuk setiap dokumen berdasarkan model Naïve Bayes yang telah diterapkan.

Tabel 4.5.1 Hasil Prediksi Naïve Bayes Classifier

No	Dokumen	Probabilitas Positif	Probabilitas Negatif	Prediksi Sentimen	Sentimen Asli
1	cemburu, suami, nekat, aniaya, istri, tewas	0.04167	0.12025	Negative	Negative
2	kaca, celaka, api, kendaraan, mogok, rel, sinau	0.04861	0.10127	Negative	Negative
3	tangki, celaka, maut, jadi, sangka	0.03472	0.07595	Negative	Negative
4	edukasi, sehat, paguyuban, gelar	0.0625	0.02532	Positive	Positive
5	kemarau, ganggu	0.01389	0.03165	Negative	Negative
6	rakit, jual, senjata, api, ilegal	0.03472	0.08861	Negative	Negative
7	harga, naik	0.02083	0.03165	Negative	Negative

8	tengkar, suami, duga, aniaya, istri, tewas	0.04167	0.11392	Negative	Negative
9	keren, rawat, budaya, pamer, keris	0.06944	0.03165	Positive	Positive
10	krisis, bersih, salur, bantu, bersih, dampak	0.04167	0.06329	Negative	Negative
11	seru, estafet, tangkap, mata, tutup	0.07639	0.03797	Positive	Positive
12	optimis, investasi, naik	0.04861	0.03165	Positive	Positive
13	aman, tani, basmi, hama, tikus	0.08333	0.03165	Positive	Positive
14	bersih, camat	0.02778	0.01899	Positive	Positive
15	nikmat, lobster, rawa, bakar, rawa, pening	0.07639	0.03797	Positive	Positive
16	depresi, pria, nekat	0.02083	0.0443	Negative	Negative
17	tangki, siaga, asap	0.02083	0.0443	Negative	Negative
18	tambang, ilegal, marak, harus, legalitas	0.03472	0.07595	Negative	Negative
19	kemarau, rawan, bakar, jual, apar, naik	0.05556	0.09494	Negative	Negative
20	emas, dukung	0.02778	0.01266	Positive	Positive
21	rekonstruksi, kasus, tewas, taksi	0.02778	0.06962	Negative	Negative
22	raih, menang, persik, manfaat, absen	0.06944	0.03165	Positive	Positive
23	tani, gotong, tangkap, tikus, sawah	0.09722	0.03797	Positive	Positive
24	gaet, investor, buat, insentif, investasi	0.07639	0.03165	Positive	Positive
25	tewas, luka, luka	0.01389	0.0443	Negative	Negative
26	racik, edar, kosmetik, ilegal, tangkap	0.04861	0.07595	Negative	Negative
27	nikmat, soto, sawah, organik, murah, sehat	0.10417	0.03797	Positive	Positive
28	keren, buat, animasi, edukasi, cegah, banjir	0.0625	0.03797	Positive	Positive
29	celaka, maut, jadi, sangka	0.02778	0.05696	Negative	Negative

30	bacok, guru, semester	0.02083	0.01899	Positive	Positive
----	-----------------------	---------	---------	----------	----------

#### 4.6. 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.

#### 4.7. 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%
80:20	24	6	93.33%	100%	88%	87%	100%

Hasil evaluasi menggunakan *confusion matrix* menunjukkan presisi kelas positif sebesar 93% pada rasio data latih 90:10 dan 85:15, serta 87% pada rasio 80:20. Presisi kelas negatif mencapai 100% pada semua rasio, menandakan akurasi sistem dalam mengidentifikasi data negatif. Recall kelas positif mencapai 100% untuk semua rasio, sementara recall kelas negatif mencapai 94% (90:10 dan 85:15) dan 88% (80:10). Dengan rasio data latih yang lebih besar, model dapat memperkaya kosa kata dan meningkatkan kemampuan menghitung probabilitas kata. Secara keseluruhan, model mencapai tingkat akurasi tinggi sebesar 96.67%, menunjukkan kinerja yang baik dalam mengklasifikasikan data positif dan negatif

#### 4.8. Pseudocode Program

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



#### 4.8.1. Inisialisasi import *library*

Pseudocode di bawah ini mencakup langkah-langkah untuk memulai penggunaan pustaka-pustaka seperti readline-sync, figlet, chalk, axios, jsdom, fs, natural, dan sastrawijs. Inisialisasi import library merupakan langkah awal untuk membangun lingkungan kerja yang diperlukan dalam pengembangan perangkat lunak dan kebutuhan untuk analisis data.

```

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

#### 4.8.2. Inisialisasi variabel global

Inisialisasi variabel global seperti incrementJudul, Judul, trainingData, testData, totalV, totalDfTrainingSentimentPositive, dan totalDfTrainingSentimentNegative diinisialisasi untuk memulai proses pengumpulan dan pengelolaan data. Penetapan nilai awal dan penggunaan struktur data seperti *array* untuk penyimpanan informasi atau data.

```

1  Inisialisasi Variabel Global:
2    -inisialisasi variabel incrementJudul = 1;
3    -inisialisasi variabel trainingData dalam array untuk menyimpan data training
4    -inisialisasi variabel testData dalam array untuk menyimpan data testing
5    -inisialisasi variabel totalV untuk menyimpan total dokumen datatraining
6    -inisialisasi variabel totalDfTrainingSentimentPositive untuk menyimpan total dokumen
7      di kelas positif
8    -inisialisasi variabel totalDfTrainingSentimentNegative untuk menyimpan total dokumen
9      di kelas negatif
10

```

Gambar 4.2. Pseudocode Inisialisasi Variabel Global

#### 4.8.3. Pembuatan fungsi `getNewsFromUrl`

Dalam fungsi `getNewsFromUrl` memaparkan fungsi asinkron yang mengumpulkan judul berita dari beberapa halaman di situs web Kompas TV.

Pengguna diminta untuk memasukkan jumlah halaman yang diinginkan, lalu sistem mengirim permintaan HTTP untuk setiap halaman. Judul berita diekstrak dari respons HTML, ditambahkan ke array Judul, dan informasi proses ditampilkan kepada pengguna. Kesalahan yang mungkin terjadi ditangkap dan dicetak sebagai pesan.

```

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

#### 4.8.4. Pembuatan fungsi getBerita

Fungsi `getBerita` pada *pseudocode* di bawah ini ditujukan untuk menampilkan judul-judul berita yang telah dikumpulkan. Setelah memberikan baris kosong di konsol, fungsi ini mengecek apakah array Judul masih kosong. Jika iya, sebuah pesan informatif akan muncul, mengingatkan pengguna bahwa judul berita belum ada, dan mereka dapat mengimpor CSV atau memasukkan berita. Jika sudah ada judul berita, fungsi ini akan menampilkan setiap judul dengan nomor urut.

```

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

#### 4.8.5. Pembuatan fungsi importFromCSV

Fungsi `importFromCSV` dengan *parameter* “file” dirancang untuk mengimpor judul berita dari file CSV ke dalam program. Pada tahap ini, pseudocode mencakup proses membaca data dari file CSV, memisahkannya menjadi baris-baris, dan mengambil judul berita dari setiap baris. Data yang diambil dibersihkan dari tanda kutip ganda dan ditambahkan ke dalam array Judul. Jika proses impor berjalan dengan sukses, sebuah pesan sukses akan ditampilkan. Dalam mengatasi potensi kesalahan, pseudocode juga mencakup tangkapan kesalahan yang akan mencetak pesan kesalahan jika terjadi masalah selama proses impor.

```

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`

#### 4.8.6. Pembuatan fungsi preprocessTextAndSaveToCSV

Fungsi `preprocessTextAndSaveToCSV` merupakan tahapan pengolahan teks dan penyimpanan ke dalam file CSV. Pada awalnya, pseudocode mengecek apakah array Judul tidak kosong. Jika tidak kosong, fungsi melakukan berbagai tahapan preprocessing seperti membaca daftar stopwords dan kata dasar, kemudian membersihkan dan menyusun token dari judul berita. Hasil preprocessing kemudian disimpan dalam file CSV 'preprocessed\_berita.csv'. Pseudocode ini juga mencakup tampilan hasil preprocessing untuk setiap judul dan pesan sukses setelah proses selesai. Jika Judul kosong, fungsi memberikan pesan bahwa judul berita belum ada dan mendorong pengguna untuk mengimpor CSV atau memasukkan berita.

```

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

#### 4.8.7. Pembuatan fungsi calculateTFIDF

Fungsi calculateTFIDF dengan *parameter* sebuah data pada skripsi ini mewakili proses perhitungan Term Frequency-Inverse Document Frequency (TF-IDF) untuk data teks yang diberikan. Pseudocode ini mencakup langkah-langkah dalam mengiterasi data, menghitung frekuensi term, document frequency, dan invers document frequency, dan terakhir menghitung nilai TF-IDF untuk setiap term pada setiap dokumen. Hasilnya adalah array tfidfData yang berisi informasi tentang TF, DF, IDF, dan TF-IDF untuk setiap term pada setiap dokumen.

```

1 - Fungsi readTrainingAndTestDataFromCSV(trainingFilename, testFilename):
2 - Coba:
3 - Inisialisasi preprocessedData, dfSentimenPositive, dfSentimenNegative
4 - Baca trainingDataRaw dari trainingFilename - Pisahkan baris menjadi trainingLines
5 - Untuk setiap baris i dalam trainingLines:
6 - Pisahkan baris menjadi line
7 - Jika panjang line lebih dari 1:
8 - Ambil teks dan sentimen dari line
9 - Tambahkan teks ke preprocessedData
10 - Tambahkan dataItem ke trainingData
11 - Proses DF untuk setiap term dalam teks
12 - Baca testDataRaw dari testFilename - Pisahkan baris menjadi testLines
13 - Untuk setiap baris i dalam testLines:
14 - Pisahkan baris menjadi line
15 - Jika panjang line lebih dari 0:
16 - Ambil teks dari line
17 - Tambahkan teks ke preprocessedData
18 - Tambahkan dataItem ke testData
19 - Gabungkan trainingData dan testData menjadi combinedData
20 - Hitung tfidfData menggunakan calculateTFIDF dengan combinedData sebagai argumen
21 - Attach TF-IDF dan inisialisasi dfSentimenPos dan dfSentimenNeg ke setiap item dalam combinedData
22 - Tulis hasil ke file CSV dengan header yang sesuai
23 - Tangkap error dan cetak pesan kesalahan
24

```

Gambar 4.7. Pseudocode Function calculateTFIDF

#### 4.8.8. Pembuatan fungsi readTrainingAndTestDataFromCSV

Fungsi readTrainingAndTestDataFromCSV dengan parameter data latih dan data uji ini bertanggung jawab untuk membaca dan memproses data training dan test dari file CSV yang diberikan. Pseudocode ini mencakup langkah-langkah

seperti membaca dan preprocess setiap baris data *training* dan *test*, menghitung Term Frequency (TF) dan Document Frequency (DF) untuk setiap term dalam teks, mengupdate data frekuensi sentimen positif dan negatif, menggabungkan data training dan test, menghitung TF-IDF menggunakan fungsi calculateTFIDF, dan menyimpan hasilnya dalam file CSV. Fungsi ini juga menghasilkan informasi seperti total dokumen di kelas positif dan di kelas negatif dalam dataset training

```

1 -Fungsi readTrainingAndTestDataFromCSV(trainingFilename, testFilename):
2   -Coba:
3     -Inisialisasi preprocessedData, dfSentimenPositive, dfSentimenNegative
4     -Baca trainingDataRaw dari trainingFilename - Pisahkan baris menjadi traininglines
5     -Untuk setiap baris i dalam traininglines:
6       -Pisahkan baris menjadi line
7       -Jika panjang line lebih dari 1:
8         -Ambil teks dan sentimen dari line
9         -Tambahkan teks ke preprocessedData
10        -Tambahkan dataItem ke trainingData
11        -Proses DF untuk setiap term dalam teks
12    -Baca testDataRaw dari testFilename - Pisahkan baris menjadi testlines
13    -Untuk setiap baris i dalam testlines:
14      -Pisahkan baris menjadi line
15      -Jika panjang line lebih dari 0:
16        -Ambil teks dari line
17        -Tambahkan teks ke preprocessedData
18        -Tambahkan dataItem ke testData
19    -Gabungkan trainingData dan testData menjadi combinedData
20    -Hitung tfidfData menggunakan calculateTFIDF dengan combinedData sebagai argumen
21    -Attach TF-IDF dan inisialisasi dfSentimenPos dan dfSentimenNeg ke setiap item dalam combinedData
22    -Tulis hasil ke file CSV dengan header yang sesuai
23    -Tangkap error dan cetak pesan kesalahan
24

```

Gambar 4.8. Pseudocode Function readTrainingAndTestDataFromCSV

#### 4.8.9. Pembuatan fungsi calculateNaiveBayesProbabilities

Fungsi calculateNaiveBayesProbabilities dengan parameter data uji ini bertanggung jawab untuk menghitung probabilitas sentimen (positif/negatif) untuk setiap dokumen dalam data uji (testData) menggunakan metode Naive Bayes. Pseudocode ini mencakup langkah-langkah seperti split teks, perhitungan probabilitas positif dan negatif untuk setiap term dalam TF-IDF, perbandingan total probabilitas untuk menentukan sentimen prediksi, dan penyimpanan hasil analisis ke dalam file CSV. Fungsi ini juga menampilkan detail perhitungan probabilitas dan sentimen prediksi untuk setiap dokumen.

```

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

#### 4.8.10. Pembuatan fungsi readSentimenResult

Fungsi readSentimenResult dengan parameter nama file ini bertanggung jawab untuk membaca hasil analisis sentimen yang disimpan dalam file CSV dengan nama filename. Pseudocode ini mencakup langkah-langkah seperti membaca isi file, memisahkan data ke dalam baris, dan menyusun hasil pembacaan ke dalam array objek yang mencakup judul, sentimen prediksi, dan sentimen sebenarnya untuk setiap dokumen. Fungsi ini berusaha memastikan pembacaan data berhasil dan mengembalikan array dataResult yang berisi hasil analisis sentimen atau mengembalikan array kosong jika terjadi kesalahan.

```

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

#### 4.8.11. Pembuatan fungsi calculateMetrics

Fungsi `calculateMetrics` dengan parameter `data`, digunakan untuk menghitung dan mengevaluasi metrik klasifikasi berdasarkan hasil analisis sentimen yang disimpan. Pseudocode ini mencakup langkah-langkah seperti menginisialisasi variabel, serta melakukan iterasi pada setiap item dalam `dataResult` untuk menghitung matriks kebingungan dan metrik evaluasi seperti akurasi, presisi, dan recall untuk kedua kelas sentimen. Hasil evaluasi metrik ditampilkan dalam bentuk confusion matrix dan informasi rinci tentang akurasi, presisi, dan recall untuk kelas positif dan negatif. Jika file analisis sentimen tidak ditemukan, fungsi akan menampilkan pesan kesalahan.

```

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

#### 4.8.12. Pembuatan fungsi menu

Fungsi `menu()` merupakan inti dari program ini yang menampilkan menu pilihan kepada pengguna. Setiap opsi menu memiliki fungsionalitasnya sendiri,

seperti pengambilan berita, penyimpanan data ke file CSV, preprocessing teks, perhitungan TF-IDF, dan evaluasi performa model menggunakan confusion matrix. Pengguna dapat memilih opsi menu yang diinginkan dan program akan menjalankan fungsionalitas sesuai dengan pilihan tersebut

```

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   - Tampilan menu utama dan hasil eksekusi menu yang dipilih oleh pengguna
16

```

Gambar 4.12. Pseudocode Function menu

## 4.9. Hasil Program

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

### 4.9.1. Tampilan Menu Program

Pada tampilan menu program, pengguna akan disuguhkan beberapa pilihan fungsi yang dirancang untuk mempermudah jalannya penelitian. Antarmuka pengguna yang intuitif akan membantu pengguna dengan mudah mengakses berbagai fitur yang ada dalam program klasifikasi sentimen.



```

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

C:\Users\User>cd/skripsi

C:\skripsi>node 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

#### 4.9.2. 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 Oudetrup
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

#### 4.9.3. 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

#### 4.9.4. 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 skr x + v
'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

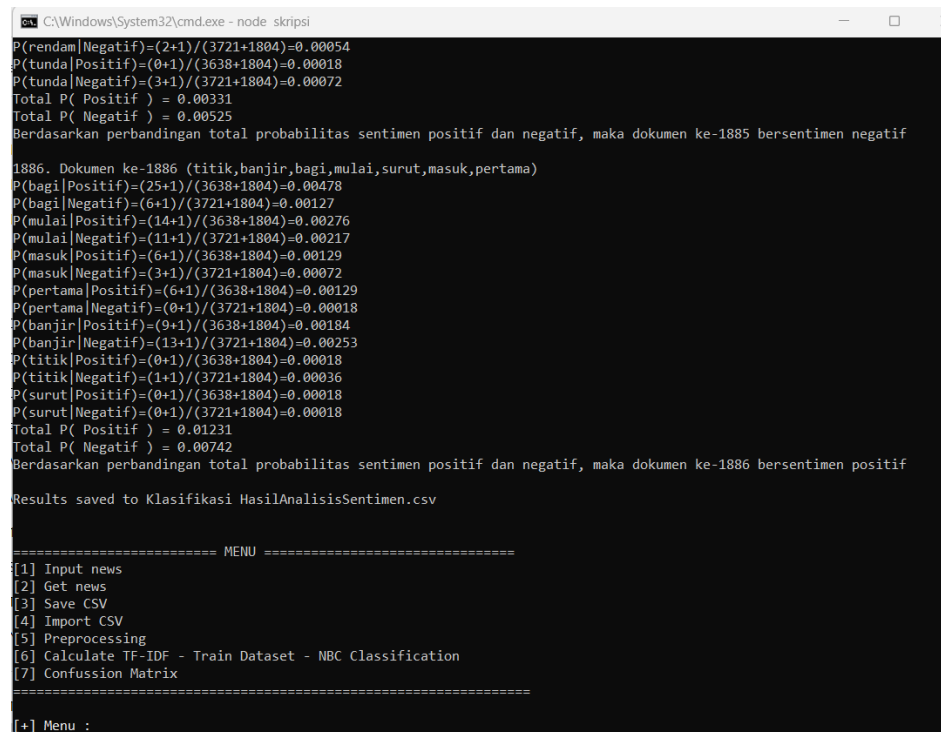
===== MENU =====
[1] Input news
[2] Get news
[3] Save CSV
[4] Import CSV
[5] Preprocessing
[6] Calculate TF-IDF - Train Dataset - NBC Classification
[7] Confussion Matrix
=====

[+] Menu :
```

Gambar 4.16. Tampilan Program *Preprocessing* Judul Berita

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

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



```

C:\Windows\System32\cmd.exe - node skripsi
P(rendam|Negatif)=(2+1)/(3721+1804)=0.00054
P(tunda|Positif)=(0+1)/(3638+1804)=0.00018
P(tunda|Negatif)=(3+1)/(3721+1804)=0.00072
Total P( Positif ) = 0.00331
Total P( Negatif ) = 0.00525
Berdasarkan perbandingan total probabilitas sentimen positif dan negatif, maka dokumen ke-1885 bersentimen negatif

1886. Dokumen ke-1886 (titik,banjir,bagi,mulai,surut,masuk,pertama)
P(bagi|Positif)=(25+1)/(3638+1804)=0.00478
P(bagi|Negatif)=(6+1)/(3721+1804)=0.00127
P(mulai|Positif)=(14+1)/(3638+1804)=0.00276
P(mulai|Negatif)=(11+1)/(3721+1804)=0.00217
P(masuk|Positif)=(6+1)/(3638+1804)=0.00129
P(masuk|Negatif)=(3+1)/(3721+1804)=0.00072
P(pertama|Positif)=(6+1)/(3638+1804)=0.00129
P(pertama|Negatif)=(0+1)/(3721+1804)=0.00018
P(banjir|Positif)=(9+1)/(3638+1804)=0.00184
P(banjir|Negatif)=(13+1)/(3721+1804)=0.00253
P(titik|Positif)=(0+1)/(3638+1804)=0.00018
P(titik|Negatif)=(1+1)/(3721+1804)=0.00036
P(surut|Positif)=(0+1)/(3638+1804)=0.00018
P(surut|Negatif)=(0+1)/(3721+1804)=0.00018
Total P( Positif ) = 0.01231
Total P( Negatif ) = 0.00742
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
[7] Confussion Matrix

[+] Menu :

```

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

#### 4.9.6. 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. Di gambar Gambar 4.18 menampilkan hasil performa metode Naïve Bayes Classifier dalam mengklasifikasikan total 1886 dataset, dengan rasio perbandingan data latih dan data uji sebesar 90:10.

```
[+] Menu : 7

-----
| Prediksi Sentimen | Sentimen Asli |
| POSITIF            | POSITIF        |
| NEGATIF            | NEGATIF        |
|-----|-----|
| POSITIF            | 841             | 70
| NEGATIF            | 96              | 879
|-----|-----|

Confusion Matrix:
- True Positives: 841
- True Negatives: 879
- False Positives: 70
- False Negatives: 96
Result:
- Akurasi: 91.20 %
- Presisi Kelas Negatif: 0.90
- Recall Kelas Negatif: 0.93
- Presisi Kelas Positif: 0.92
- Recall Kelas Positif: 0.90

===== 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
=====
```

Gambar 4.18. Tampilan Program Perhitungan Confussion Matrix

## **BAB V PENUTUP**

### **5.1. Kesimpulan**

Hasil penelitian mengenai klasifikasi sentimen pada judul berita Kompas.TV Kota Semarang menunjukkan bahwa penggunaan algoritma Naïve Bayes Classifier menghasilkan performa yang cukup baik. Dengan pembagian rasio data latih dan data uji sebesar 90:10, dengan total dataset sebesar 1886 mendapatkan akurasi mencapai tingkat 91.20%, menandakan kemampuan yang tinggi dalam mengenali sentimen pada judul berita. Lebih lanjut, ketika fokus pada kelas negatif, presisi sebesar 90% dan recall sebesar 93% menunjukkan ketepatan dalam mengidentifikasi judul berita dengan sentimen negatif.

Di sisi lain, untuk kelas positif, metode Naïve Bayes Classifier juga menunjukkan hasil yang positif dengan presisi sebesar 92% dan recall sebesar 90%. Artinya, algoritma ini efektif dalam mengklasifikasikan judul berita yang membawa sentimen positif. Secara keseluruhan, temuan ini memberikan kontribusi positif dalam memahami analisis sentimen pada judul berita di platform media online, khususnya dalam konteks Kompas.TV di Kota Semarang.

### **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. Untuk meningkatkan kualitas penelitian dan memberikan kontribusi lebih lanjut, berikut adalah beberapa saran yang diusulkan oleh peneliti:

1. Bagi peneliti yang akan datang dilakukan evaluasi mendalam terhadap kamus stopwords yang digunakan pada proses stopwords removal. Peneliti masa depan sebaiknya memastikan bahwa kata-kata yang tidak memiliki makna atau sentimen yang signifikan dapat terfilter secara efektif. Mungkin mempertimbangkan penggunaan kamus stopwords yang lebih spesifik atau pendekatan kontekstual dapat meningkatkan kualitas pengolahan data.

2. Bagi peneliti yang akan datang dengan topik pengklasifikasian sentimen judul berita, disarankan untuk tidak hanya terpaku pada metode Naive Bayes Classifier. Eksplorasi algoritma lain seperti Support Vector Machines (SVM), Decision Trees, atau Deep Learning dapat memberikan wawasan tambahan dan meningkatkan performa klasifikasi sentimen. Diversifikasi metode dapat memberikan pemahaman lebih baik tentang kecocokan algoritma dengan data yang digunakan.

## Daftar Pustaka

Anton Mabururi Kn. (2018). *Drama Produksi Program Tv : Manajemen Produksi Dan Penulisan Naskah / Anton Mabururi Kn.* Gramedia Pustaka Utama.

Asfi, M., & Fitrianiingsih, N. (2020). Implementasi Algoritma Naive Bayes Classifier Sebagai Sistem Rekomendasi Pembimbing Skripsi. *Infotekjar Jurnal Nasional Informatika Dan Teknologi Jaringan*, 5(1), 44.

Biro Humas Kementerian Kominfo. (2023). *Kembangkan Ekonomi Digital, Indonesia Butuh Talenta Digital Berkualitas.*  
[https://www.kominfo.go.id/content/detail/47689/Siaran-Pers-No-28hmkominfo032023-Tentang-Kembangkan-Ekonomi-Digital-Indonesia-Butuh-Talenta-Digital-Berkualitas/0/Siaran\\_Pers](https://www.kominfo.go.id/content/detail/47689/Siaran-Pers-No-28hmkominfo032023-Tentang-Kembangkan-Ekonomi-Digital-Indonesia-Butuh-Talenta-Digital-Berkualitas/0/Siaran_Pers)

Delfariyadi, F., Helen, A., & Yulawati, S. (2022). Klasifikasi Sentimen Judul Berita Pemberitaan Covid-19 Tahun 2021 Pada Media Detikhealth. *Journal Of Information Engineering And Educational Technology*, 6(2), 50–57.  
<https://doi.org/10.26740/jieet.v6n2.p50-57>

Ferryawan, R., Kusriani, K., & Wibowo, F. W. (2020). Analisis Sentimen Wisata Jawa Tengah Menggunakan Naïve Bayes. *Jurnal Informa : Jurnal Penelitian Dan Pengabdian Masyarakat*, 5(3), 55–60.  
<https://doi.org/10.46808/informa.v5i3.146>

Jaya, C., & Muslim, K. (2019). Analisis Sentimen Pada Media Daring Tentang Pemilihan Presiden Indonesia Tahun 2019 Menggunakan Metode Naïve Bayes. *E-Proceeding Of Engineering*, 6(2), 9698.

Keputusan Dirjen Penguatan Riset Dan Pengembangan Ristek Dikti, S., Kurnia Sari, W., Palupi Rini, D., Firsandaya Malik, R., & Saladin Azhar, I. B. (2017). Terakreditasi Sinta Peringkat 2 Klasifikasi Teks Multilabel Pada Artikel Berita Menggunakan Long Short-Term Memory Dengan Word2vec. *Masa Berlaku Mulai*, 1(3), 276–285.

Kurniawan, I., Studi, P., Departemen, S., Matematika, F., Ilmu, D. A. N., Alam, P., & Hasanuddin, U. (2022). *Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Sains Pada Program Studi Statistika Departemen Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Hasanuddin Iwan Kurniawan*.

Kusnia, U. (2022). *ii ii Sentiment Analysis Of News App Reviews Online On Google Play Using Naive Bayes Classifier Algorithm And Support Vector Machines Methods Thesis By: Ulfa Kusnia Nim. 19841003 Program Studi Magister Informatika Fakultas Sains Dan Teknologi Universitas*.

Miftakhurahmat, M. A., Safitri, N., Kusnadi, P. A., & Rozikin, C. (2023). Klasifikasi Pengguna Hashtag Pada Aplikasi Tiktok Menggunakan Perbandingan Metode K-Nearest Neighbors Dan Naïve Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3), 427–433.  
<https://doi.org/10.23960/Jitet.V11i3.3150>

Provost, F., & Fawcett, T. (2019). *Data Science For Business*. O'reilly Media.

Rama, P., Putra, B., & Perdana, R. S. (2023). *Klasifikasi Judul Berita Online Menggunakan Metode Support Vector Machine ( Svm ) Dengan Seleksi Fitur Chi-*



*Square*. 7(5), 2132–2141.

Roufia, A. (2018). *Text Mining Dengan Metode Naïve Bayes Classifier Untuk Mengklasifikan Berita Berdasarkan Konten*. 8.

Sholih 'Afif, M., Muzakir, M., Al, M. I., & Al Awalaïen, G. (2021). Text Mining Untuk Mengklasifikasi Judul Berita Online Studi Kasus Radar Banjarmasin Menggunakan Metode Naïve Bayes. *Kumpulan Jurnal Ilmu Komputer (Klik)*, 08(2), 199–208.

Silge, J., & Robinson, D. (2019). *Text Mining With R: A Tidy Approach*. O'reilly Media. <https://doi.org/10.1002/9781119447753>

Singh, G., Yadav, B. K., Singh, B. P., & Sharama, S. (2022). Identification Of Fake News Using Machine Learning Approach. *Proceedings - 2022 4th International Conference On Advances In Computing, Communication Control And Networking, Icac3n 2022*, 271–274. <https://doi.org/10.1109/Icac3n56670.2022.10074374>

Umi Rofiqoh<sup>1</sup>, Rizal Setya Perdana<sup>2</sup>, M. A. F., & Program. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine Dan Lexicon Based Features Twitter Event Detection View Project Human Detection And Tracking View Project. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12)(October), 1725–1732. <https://www.researchgate.net/publication/320234928>

## LAMPIRAN

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

$$P(\text{cemburu}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{cemburu}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{suami}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{suami}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{nekat}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{nekat}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{aniaya}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{aniaya}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{istri}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{istri}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{tewas}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{tewas}|\text{Negatif}) = \frac{4 + 1}{66 + 92} = 0.03165$$

$$\text{Total } P(\text{Positif}) = 0.04167$$

$$\text{Total } P(\text{Negatif}) = 0.12025$$

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)

$$P(\text{kaca}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{kaca}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{celaka}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{celaka}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{api}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{api}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{kendaraan}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{kendaraan}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{mogok}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{mogok}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{rel}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{rel}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{sinau}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{sinau}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$\text{Total } P(\text{Positif}) = 0.04861$$

$$\text{Total } P(\text{Negatif}) = 0.10127$$

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

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

$$P(\text{celaka}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{celaka}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{tangki}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{tangki}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{maut}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{maut}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{jadi}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{jadi}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{sangka}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{sangka}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$\text{Total } P(\text{Positif}) = 0.03472$$

$$\text{Total P( Negatif )} = 0.07595$$

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

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

$$P(\text{edukasi}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{edukasi}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{sehat}|\text{Positif}) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(\text{sehat}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{paguyuban}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{paguyuban}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{gelar}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{gelar}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$\text{Total P( Positif )} = 0.06250$$

$$\text{Total P( Negatif )} = 0.02532$$

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

5. Dokumen ke-5 (kemarau, ganggu)

$$P(\text{kemarau}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{kemarau}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{ganggu}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{ganggu}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$\text{Total } P(\text{Positif}) = 0.01389$$

$$\text{Total } P(\text{Negatif}) = 0.03165$$

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

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

$$P(\text{api}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{api}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{rakit}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{rakit}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{jual}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{jual}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{senjata}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{senjata}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{ilegal}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{ilegal}|\text{Negatif}) = \frac{3 + 1}{66 + 92} = 0.02532$$

$$\text{Total } P(\text{Positif}) = 0.03472$$

$$\text{Total } P(\text{Negatif}) = 0.08861$$

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

#### 7. Dokumen ke-7 (harga, naik)

$$P(\text{harga}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{harga}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{naik}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{naik}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$\text{Total } P(\text{Positif}) = 0.02083$$

$$\text{Total } P(\text{Negatif}) = 0.03165$$

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

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

$$P(\text{suami}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{suami}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{aniaya}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{aniaya}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{istri}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{istri}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{tewas}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{tewas}|\text{Negatif}) = \frac{4 + 1}{66 + 92} = 0.03165$$

$$P(\text{tengkar}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{tengkar}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{duga}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{duga}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$\text{Total } P(\text{Positif}) = 0.04167$$

$$\text{Total } P(\text{Negatif}) = 0.11392$$

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

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

$$P(\text{keren}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{keren}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{rawat}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{rawat}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$



$$P(\text{budaya}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{budaya}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{pamer}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{pamer}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{keris}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{keris}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$\text{Total } P(\text{Positif}) = 0.06944$$

$$\text{Total } P(\text{Negatif}) = 0.03165$$

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

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

$$P(\text{krisis}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{krisis}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{bersih}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{bersih}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{salur}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{salur}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{bantu}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{bantu}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{dampak}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{dampak}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$\text{Total } P(\text{Positif}) = 0.04167$$

$$\text{Total } P(\text{Negatif}) = 0.06329$$

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

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

$$P(\text{seru}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{seru}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{estafet}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{estafet}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{tangkap}|\text{Positif}) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(\text{tangkap}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{mata}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{mata}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{tutup}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{tutup}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$\text{Total } P(\text{Positif}) = 0.07639$$

$$\text{Total } P(\text{Negatif}) = 0.03797$$

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

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

$$P(\text{naik}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{naik}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{optimis}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{optimis}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{investasi}|\text{Positif}) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(\text{investasi}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$\text{Total } P(\text{Positif}) = 0.04861$$

$$\text{Total } P(\text{Negatif}) = 0.03165$$

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

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

$$P(\text{aman}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{aman}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{tani}|\text{Positif}) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(\text{tani}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{basmi}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{basmi}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{hama}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{hama}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{tikus}|\text{Positif}) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(\text{tikus}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$\text{Total } P(\text{Positif}) = 0.08333$$

$$\text{Total } P(\text{Negatif}) = 0.03165$$

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

14. Dokumen ke-14 (bersih, camat)

$$P(\text{bersih}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{bersih}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{camat}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{camat}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$\text{Total } P(\text{Positif}) = 0.02778$$

$$\text{Total } P(\text{Negatif}) = 0.01899$$

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

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

$$P(\text{nikmat}|\text{Positif}) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(\text{nikmat}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{lobster}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{lobster}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{rawa}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{rawa}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{bakar}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{bakar}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{pening}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{pening}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$Total P( Positif ) = 0.07639$$

$$Total P( Negatif ) = 0.03797$$

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

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

$$P(nekat|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(nekat|Negatif) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(depresi|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(depresi|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(pria|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(pria|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$Total P( Positif ) = 0.02083$$

$$Total P( Negatif ) = 0.04430$$

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

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

$$P(tangki|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(tangki|Negatif) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(siaga|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(siaga|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(asap|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(asap|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$Total P( Positif ) = 0.02083$$

$$Total P( Negatif ) = 0.04430$$

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

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

$$P(ilegal|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(ilegal|Negatif) = \frac{3 + 1}{66 + 92} = 0.02532$$

$$P(tambang|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(tambang|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(marak|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(marak|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(harus|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(harus|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(legalitas|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{legalitas}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$\text{Total } P(\text{Positif}) = 0.03472$$

$$\text{Total } P(\text{Negatif}) = 0.07595$$

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

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

$$P(\text{kemarau}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{kemarau}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{jual}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{jual}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{naik}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{naik}|\text{Negatif}) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(\text{bakar}|\text{Positif}) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(\text{bakar}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{rawan}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{rawan}|\text{Negatif}) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(\text{apar}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$



$$P(apar|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$Total P( Positif ) = 0.05556$$

$$Total P( Negatif ) = 0.09494$$

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

20. Dokumen ke-20 (emas, dukung)

$$P(emas|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(emas|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(dukung|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(dukung|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$Total P( Positif ) = 0.02778$$

$$Total P( Negatif ) = 0.01266$$

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

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

$$P(tewas|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(tewas|Negatif) = \frac{4 + 1}{66 + 92} = 0.03165$$

$$P(rekonstruksi|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(rekonstruksi|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(kasus|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(kasus|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(taksi|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(taksi|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$Total P( Positif ) = 0.02778$$

$$Total P( Negatif ) = 0.06962$$

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

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

$$P(raih|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(raih|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(menang|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(menang|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(persik|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(persik|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(manfaat|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(manfaat|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(absen|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(absen|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$Total P( Positif ) = 0.06944$$

$$Total P( Negatif ) = 0.03165$$

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

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

$$P(tangkap|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(tangkap|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(tani|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(tani|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(tikus|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(tikus|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(gotong|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(gotong|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(sawah|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(sawah|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$Total P( Positif ) = 0.09722$$

$$Total P( Negatif ) = 0.03797$$

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

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

$$P(investasi|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(investasi|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(gaet|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(gaet|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(investor|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(investor|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(buat|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(buat|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(insentif|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(insentif|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$Total P( Positif ) = 0.07639$$

$$Total P( Negatif ) = 0.03165$$

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

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

$$P(tewas|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(tewas|Negatif) = \frac{4 + 1}{66 + 92} = 0.03165$$

$$P(luka|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(luka|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$Total P( Positif ) = 0.01389$$

$$Total P( Negatif ) = 0.04430$$

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

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

$$P(ilegal|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(ilegal|Negatif) = \frac{3 + 1}{66 + 92} = 0.02532$$

$$P(tangkap|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(tangkap|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(racik|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(racik|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(edar|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(edar|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(kosmetik|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(kosmetik|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$Total P( Positif ) = 0.04861$$

$$Total P( Negatif ) = 0.07595$$

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

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

$$P(sehat|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(sehat|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(nikmat|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(nikmat|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(sawah|Positif) = \frac{2 + 1}{52 + 92} = 0.02083$$

$$P(sawah|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(soto|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(soto|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(organik|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(organik|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(murah|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(murah|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$Total P( Positif ) = 0.10417$$

$$Total P( Negatif ) = 0.03797$$

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

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

$$P(edukasi|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(edukasi|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(keren|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(keren|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(buat|Positif) = \frac{1 + 1}{52 + 92} = 0.01389$$

$$P(buat|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(animasi|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(animasi|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(cegah|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(cegah|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(banjir|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(banjir|Negatif) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$Total P( Positif ) = 0.06250$$

$$Total P( Negatif ) = 0.03797$$

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

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

$$P(celaka|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(celaka|Negatif) = \frac{2 + 1}{66 + 92} = 0.01899$$

$$P(maut|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(maut|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(jadi|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(jadi|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$

$$P(sangka|Positif) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(sangka|Negatif) = \frac{1 + 1}{66 + 92} = 0.01266$$



$$\text{Total } P(\text{Positif}) = 0.02778$$

$$\text{Total } P(\text{Negatif}) = 0.05696$$

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

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

$$P(\text{bacok}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{bacok}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{guru}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{guru}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$P(\text{semester}|\text{Positif}) = \frac{0 + 1}{52 + 92} = 0.00694$$

$$P(\text{semester}|\text{Negatif}) = \frac{0 + 1}{66 + 92} = 0.00633$$

$$\text{Total } P(\text{Positif}) = 0.02083$$

$$\text{Total } P(\text{Negatif}) = 0.01899$$

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