



Penerapan Metode *Naïve Bayes Classifier* untuk Klasifikasi Sentimen pada Judul Berita

Alrico Rizki Wibowo¹, Yani Parti Astuti^{*2}, Etika Kartikadarma³, Egia Rosi Subhiyakto⁴, Nurul Anisa Sri Winarsih⁵, Muhammad Syaifur Rohman⁶

^{1,2,3,4,5,6}Universitas Dian Nuswantoro, Semarang, Indonesia

E-mail : ¹111202012703@mhs.dinus.ac.id, ^{2*}yanipartiastuti@dsn.dinus.ac.id,

³etika.kartikadarma@dsn.dinus.ac.id, ⁴egia@dsn.dinus.ac.id, ⁵nurulanisasw@dsn.dinus.ac.id,

⁶syaifur@dsn.dinus.ac.id

**Penulis Korespondensi*

Received 25 May 2024; Revised 13 June 2024; Accepted 19 June 2024

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.

Keywords: Kompas.TV, Analisis Sentimen, Naïve Bayes Classifier.

Abstract - News has a major role as a source of information to convey reports on opinions, events, and the latest findings in various aspects of life. News headlines, as an important component, can be a determinant of news content. The sentiment contained in news headlines can be classified using sentiment analysts, as is the case in the online media platform Kompas.TV. News headlines are retrieved using an automated program that utilises the HTML body with the help of NodeJs as the technology for program creation. This research is focused on the application of Naïve Bayes Classifier method to classify sentiment on Kompas.TV news headlines in Semarang City. The results showed an accuracy rate of 91.04%, with a ratio of training data and test data of 90:10. The conclusion of this study is that the Naïve Bayes Classifier method is effective in identifying news headlines with negative sentiment on Kompas.TV, with a precision of 89% and recall of 94%. This finding makes a positive contribution to the understanding of sentiment analysis on news headlines in online media, especially in the context of Kompas.TV news in Semarang City.

Keywords: Kompas.TV, Sentiment Analysis, Naïve Bayes Classifier.

1. PENDAHULUAN

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. Dalam berita pasti adanya sebuah judul yang merupakan bagian atas atau kepala yang akan menjadi informasi utama untuk pembaca (Delfariyadi et al., 2022). Berita dulu hanya muncul di surat kabar, radio, dan televisi. Seiring dengan kemajuan teknologi dan peningkatan jumlah pengguna internet di Indonesia, maka sumber informasi berita kini lebih dominan berasal dari media *online*. Media *online* seperti berita online, memiliki unsur-unsur seperti *headline*, *lead*, tubuh berita, dan foto (Dewi et al., 2024). 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 (Delfariyadi et al., 2022).

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 (Miftakhurahmat et al., 2023). 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 (Mulyani & Novita, 2022) (Sholih & Muzakir, 2021). 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 sentiment (Ruus et al., 2022) (Rofqoh et al., 2017). Menurut penelitian (Delfariyadi et al., 2022), sentimen dapat diklasifikasikan menggunakan metode *Naïve Bayes Classifier*. Metode *Naïve Bayes Classifier* telah terbukti efektif dalam pengklasifikasian sentimen dengan kecepatan pemrosesan dan pengklasifikasian data yang tinggi. 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 *Naïve 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.

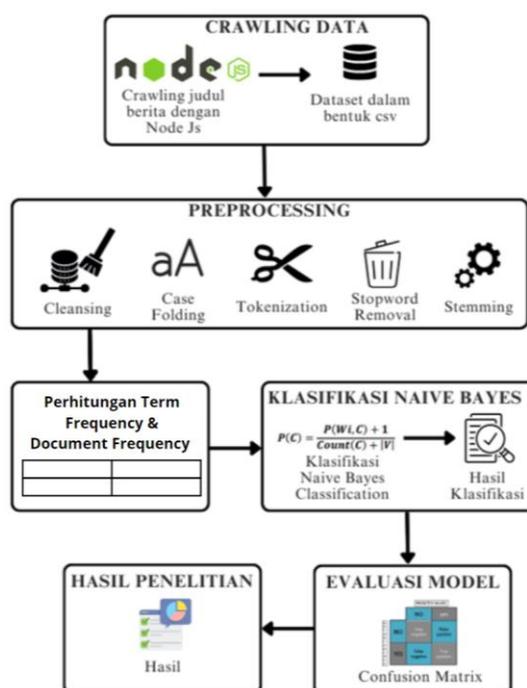


2. METODE PENELITIAN

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 (Tan et al., 2023). 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 ("Identification of Fake News Using Machine Learning Approach," n.d.) (Bahtiar et al., 2023).

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 (Fay, 2018).



Gambar 1 Alur Penelitian

2.1 Crawling Data

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 Pelabelan Dataset

Proses pelabelan dataset akan dilakukan secara manual oleh seorang penganalisis yang memiliki latar belakang gelar sarjana sastra. Langkah ini penting untuk mengukur hasil prediksi dari program yang akan dikembangkan menggunakan metode pengklasifikasi *Naïve Bayes Classifier*. Proses pelabelan dilakukan dengan teliti untuk memastikan setiap data dalam dataset diberikan label sentimen yang sesuai berdasarkan pemahaman mendalam terhadap struktur bahasa dan konteks budaya Indonesia

2.3 Text Preprocessing

Preprocessing dalam 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:

- Cleansing*, merupakan tahap pembersihan data dilakukan untuk menghilangkan karakter atau informasi yang tidak relevan atau mengganggu dalam judul berita.
- Case Folding*, merupakan tahap dalam *preprocessing* yang dilakukan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil semua.
- 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.
- 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.
- 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, -l, dll dan dilakukan pencocokan kata dasar didalam kamus Bahasa Indonesia.

2.4 Term Frequency & Document Frequency

Term Frequency mengukur seberapa sering suatu kata muncul dalam suatu dokumen sesuai pada kelasnya, sementara *Document Frequency* mengukur berapa banyak dokumen yang mengandung kata tersebut sesuai pada kelasnya. Dalam perhitungan TF dan DF berguna untuk menghitung probabilitas dalam pengklasifikasian, berikut adalah rumus untuk mencari TF dan DF

$$TF = TF(w, d) \quad (1)$$

Keterangan :

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

$$DF = DF(w) \quad (2)$$

Keterangan :

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

2.5 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 & Fitriarningsih, 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
Count(C) : Jumlah semua *terms* yang ada di dokumen data latih pada kelas C
|V| : Jumlah total *terms* (*term* yang sama dihitung 1)

2.6 Confusion Matrix

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

a. Accuracy

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

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

Keterangan:

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

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

c. *Recall*

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

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

Keterangan:

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

Tabel 1 *Confussion Matrix*

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

3. HASIL DAN PEMBAHASAN

3.1 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 2 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 Diredam Banjir, Perjalanan Kereta Ditunda!
1886	Ada 30 Titik Banjir di Semarang, Sebagian Mulai Surut Masuk Hari Pertama Tahun 2023



3.2 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 berdasarkan kata yang memiliki bobot sentimen paling besar. Berikut adalah contoh dataset yang telah dilakukan pelabelan :

Tabel 3 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 Diredam Banjir, Perjalanan Kereta Ditunda!	Negative
1886	Ada 30 Titik Banjir di Semarang, Sebagian Mulai Surut Masuk Hari Pertama Tahun 2023	Positive

3.3 Text Preprocessing

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* (Aditomo Mahardika Putra, 2021) (Maulana et al., 2024).

Tabel 3 Hasil Text Preprocessing

No	Dokumen
1	cemburu,suami,nekat,aniaya,istri,tewas
2	kaca,celaka,api,kendara,mogok,rel,sinau
3	tangki,celaka,maut,jadi,sangka
...	...
28	keren,buat,animasi,edukasi,cegah,banjir
29	celaka,maut,jadi,sangka
30	bacok,guru,semester



Tabel 4 Hasil Sampel Data Training

No	Dokumen	Sentimen
1	cemburu,suami,nekat,aniaya,istri,tewas	Negative
2	kaca,celaka,api,kendara,mogok,rel,sinau	Negative
3	tangki,celaka,maut,jadi,sangka	Negative
...
25	tewas,luka,luka	Negative
26	racik,edar,kosmetik,ilegal,tangkap	Negative
27	nikmat,soto,sawah,organik,murah,sehat	Positive

Tabel 5 Hasil Sampel Data Uji

No	Dokumen
1	keren,buat,animasi,edukasi,cegah,banjir
2	celaka,maut,jadi,sangka
3	bacok,guru,semester

3.4 Perhitungan Term Frequency & Document Frequency

Tahap selanjutnya dalam proses adalah perhitungan *Term Frequency*, dimana term dihitung berdasarkan kemunculan kata pada dokumen dan menampilkan total terms dokumen berdasarkan kelasnya. Berikut merupakan dataset sampel untuk *training* dan *testing* yang sudah melalui proses pembobotan kata:

Tabel 6 Hasil Term Frequency Sampel Dataset Training

No	Term	T F	D F	Total Terms di Dokumen Positif	Total Terms di Dokumen Negatif
1	cemburu	1	1	0	1
2	suami	1	2	0	2
3	nekat	1	2	0	2
...
90	soto	1	1	1	0
91	organik	1	1	1	0
92	murah	1	1	1	0
TOTAL					
TOTAL DOKUMEN DI KELAS POSITIF					52
TOTAL DOKUMEN DI KELAS NEGATIF					65
TOTAL TERMS DATA LATIH					92

3.5 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:

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

3.5.1 Proses Perhitungan Probabilitas Data Latih

Dalam Tabel 6 Hasil TF & DF Sampel Dataset *Training*, diperoleh: $\text{Count}(\text{positif}) = 52$, $\text{Count}(\text{negatif}) = 65$, dan $|V| = 92$.

Maka didapatkan probabilitas prior untuk setiap kelas C:

Dengan rumus,

$$P(C) = \frac{\text{Count}(C)}{|V|}$$

$$P(\text{Positif}) = \frac{52}{92} = 0.56522$$

$$P(\text{Negatif}) = \frac{66}{92} = 0.71739$$

Dilanjutkan menghitung probabilitas kondisional untuk setiap *term* atau kata dalam dokumen terhadap kelas positif dan negatif, dengan perhitungan berikut ini:

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

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

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

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

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

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

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

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

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

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

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

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

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

Setelah dilakukan perhitungan probabilitas kondisional, selanjutnya mengalikan semua probabilitas *terms* dengan probabilitas prior

$$P(\text{Positif}|\text{Dokumen 1}) = P(\text{Positif}) * P(\text{cemburu}|\text{Positif}) * P(\text{suami}|\text{Positif}) * P(\text{nekat}|\text{Positif}) * P(\text{aniaya}|\text{Positif}) * P(\text{istri}|\text{Positif}) * P(\text{tewas}|\text{Positif})$$

$$P(\text{Negatif}|\text{Dokumen 1}) = P(\text{Negatif}) * P(\text{cemburu}|\text{Negatif}) * P(\text{suami}|\text{Negatif}) * P(\text{nekat}|\text{Negatif}) * P(\text{aniaya}|\text{Negatif}) * P(\text{istri}|\text{Negatif}) * P(\text{tewas}|\text{Negatif})$$

$$P(\text{Positif}|\text{Dokumen 1}) = 0.56521 * 0.00694 * 0.00694 * 0.00694 * 0.00694 * 0.00694 * 0.00694$$

$$P(\text{Positif}|\text{Dokumen 1}) = 6.33929e-14$$



$$P(\text{Negatif} | \text{Dokumen 1}) = 0.71739 * 0.01266 * 0.01899 * 0.01899 * * 0.01899 * * 0.01899 * 0.03165$$

$$P(\text{Negatif} | \text{Dokumen 1}) = 3.73507e-11$$

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

3.5.2 Proses Perhitungan Probabilitas Data Uji

Tahap selanjutnya dari penelitian ini adalah perhitungan dengan menggunakan data uji berdasarkan Hasil *Term Frequency* Sampel Dataset Training. Setelah melalui seluruh perhitungan pada data uji, hasil dari perhitungan probabilitas sentimen dari setiap dokumen dikumpulkan dan disajikan dalam Tabel 7, 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 7 Hasil Prediksi Naïve Bayes Classifier

No	Dokumen	Probabilitas Positif	Probabilitas Negatif	Prediksi Sentimen	Sentimen Asli
1	cemburu, suami, nekat, aniaya, istri, tewas	6.31E-14	3.74E-11	Negative	Negative
2	kaca, celaka, api, kendaraan, mogok, rel, sinau	4.40E-16	8.41E-14	Negative	Negative
3	tangki, celaka, maut, jadi, sangka	9.13E-12	5.25E-10	Negative	Negative
...
28	keren, buat, animasi, edukasi, cegah, banjir	5.07E-13	4.61E-14	Positive	Positive
29	celaka, maut, jadi, sangka	1.31E-09	2.76E-08	Negative	Negative
30	bacok, guru, semester	1.89E-07	1.82E-07	Positive	Positive

3.6 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 *confussion matrix*:

Tabel 8 Hasil *Confussion Matrix*

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



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

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

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

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

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

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

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

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. 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 94.64%, menandakan kemampuan yang tinggi dalam mengenali sentimen pada judul berita. Lebih lanjut, ketika fokus pada kelas negatif, presisi sebesar 94% dan *recall* sebesar 95% 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 95% dan *recall* sebesar 94%. 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.

DAFTAR PUSTAKA

- Aditomo Mahardika Putra, R. (2021). Underground Support System Determination: A Literature Review. *International Journal of Research Publications*, 83(1). <https://doi.org/10.47119/IJRP100831820212185>
- 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.
- Bahtiar, S. A. H., Dewa, C. K., & Luthfi, A. (2023). Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling. *Journal of Information Systems and Informatics*, 5(3), 915–927. <https://doi.org/10.51519/journalisi.v5i3.539>
- Delfariyadi, F., Helen, A., & Yuliawati, 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>



- Dewi, N. L. P. R., Wijaya, I. N. S. W., Purnamawan, I. K., & Marti, N. W. (2024). Model *Classifier* Judul Berita Pariwisata Indonesia Berdasarkan Sentimen. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 117–124. <https://doi.org/10.25126/jtiik.20241117617>
- Fay, C. (2018). Text Mining with *R*: A Tidy Approach. *Journal of Statistical Software*, 83(Book Review 1). <https://doi.org/10.18637/jss.v083.b01>
- Identification of Fake News Using Machine Learning Approach. (n.d.). *IEEE Explore*.
- Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM): Sentiment Analysis of Pluang Applications With Naive Bayes and Support Vector Machine (SVM) Algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 375–384. <https://doi.org/10.57152/malcom.v4i2.1206>
- Miftakurahmat, Moh. 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). <https://doi.org/10.23960/jitet.v11i3.3150>
- Mulyani, S., & Novita, R. (2022). IMPLEMENTATION OF THE NAIVE BAYES CLASSIFIER ALGORITHM FOR CLASSIFICATION OF COMMUNITY SENTIMENT ABOUT DEPRESSION ON YOUTUBE. *Jurnal Teknik Informatika (Jutif)*, 3(5), 1355–1361. <https://doi.org/10.20884/1.jutif.2022.3.5.374>
- Rofqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). *Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features*.
- Roufia, A. (2018). *Text Mining Dengan Metode Naive Bayes Classifier Untuk Mengklasifikasi Berita Berdasarkan Konten*. 8.
- Ruus, E. G., Latumakulita, L. A., & Prang, J. D. (2022). *Analisis Sentimen di Media Online menggunakan Metode Naive Bayes*. 11(1).
- Sholih, M., & Muzakir, M. (2021). *TEXT MINING UNTUK MENGLASIFIKASI JUDUL BERITA ONLINE STUDI KASUS RADAR BANJARMASIN MENGGUNAKAN METODE NAÏVE BAYES*. 08(2).
- Singh, G., Yadav, B. K., Singh, B. P., & Sharama, S. (2022). Identification of Fake News Using Machine Learning Approach. *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 271–274. <https://doi.org/10.1109/ICAC3N56670.2022.10074374>
- Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences*, 13(7), 4550. <https://doi.org/10.3390/app13074550>