Jurnal Elektronik Ilmu Komputer Udayana
Volume 14, No 2. November 2025

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Analisis Sentimen Review Toko Menggunakan Naive Bayes dengan Penanganan Negasi dan Mutual Information

Monika Hermiani Yolanda Simamora[a1], I Gede Surya Rahayuda[a2], Ngurah Agus Sanjaya ER[a3], I Gusti Ngurah Anom Cahyadi Putra[a4]

[a]Informatics Departement, Faculty of Mathematics and Natural Science, Udayana University
South Kuta, Badung, Bali, Indonesia
[1]monikasimamora8@email.com
[2]igedesuryarahayuda@unud.ac.id
[3]agus_sanjaya@unud.ac.id
[4]anom.cp@unud.ac.id

Abstract

Sentiment analysis refers to a technique within natural language processing for extracting and interpreting sentiments or opinions regarding a particular topic, such as product reviews. The model in this study was built through a series of stages, including preprocessing, feature extraction using TF-IDF, feature selection using Mutual Information, and sentiment classification with Multinomial Naive Bayes algorithm. The preprocessing stage consists of several processes, namely cleaning, case folding, tokenization, normalization, stopwords removal, stemming, and handling of negations and intensifiers. This study was tested using four scenarios. In the first scenario, classification was performed using the Multinomial Naive Bayes algorithm as a baseline approach. The second scenario involved classification with the handling of negations and intensifiers during the preprocessing stage. The third scenario involved classification with feature selection using Mutual Information. The fourth scenario combined the handling of negations and intensifiers with feature selection using Mutual Information. Based on the results, the classification using Multinomial Naive Bayes with the combination of negation and intensifier handling and feature selection using Mutual Information delivered the highest performance, attaining 74.08% accuracy, 74.32% precision, 74.19% recall, and an F1-Score of 74.04% with 60% of the features selected.

Keywords: *Sentiment Analysis, Negation Handling, TF-IDF, Mutual Information, Multinomial Naive Bayes*

1. Pendahuluan

Analisis sentimen adalah salah satu bidang dalam pemrosesan bahasa alami yang dapat mengidentifikasi dan memahami data tekstual yang berupa opini atau sentimen terkait suatu topik kemudian mengklasifikasikannya menjadi kelas sentimen positif, netral, atau negatif [1]. Salah satu algoritma populer dalam analisis sentimen adalah Multinomial Naive Bayes. Multinomial Naive Bayes adalah algoritma klasifikasi yang menggunakan prinsip perhitungan probabilitas [2]. Cara kerja dari algoritma ini yaitu menghitung frekuensi kemunculan setiap kata pada dokumen secara independen [3]. Mandasari dkk dalam penelitiannya membahas tentang analisis sentimen dengan algoritma Multinomial Naive Bayes dan memperoleh nilai akurasi 86,57% [4].

Metode Multinomial Naive Bayes yang menghitung frekuensi setiap kata secara independen ini menghadapi masalah ketika adanya kalimat yang mengandung kata negasi atau kata yang membalik makna pesan. Misalnya, kata “tidak bagus” yang bermakna negatif namun dapat diklasifikasikan sebagai sentimen positif karena memperhatikan kata “bagus” yang bermakna positif. Oleh sebab itu, penanganan negasi kata diperlukan dalam proses analisis sentimen [5]. Penanganan negasi ini telah diterapkan pada penelitian yang dilakukan oleh Ananda dkk. Berdasarkan pene litian tersebut, diketahui bahwa penerapan penanganan negasi dapat meningkatkan nilai akurasi dari 83,7% menjadi 84% [6]. Adapun penelitian sebelumnya yang dilakukan oleh Darmawan dkk menghasilkan akurasi rata-rata sebesar 80,514% saat mempertahankan negasi lebih tinggi dibandingkan saat menghilangkan negasi dengan nilai akurasi rata-rata sebesar 79,348% [7].

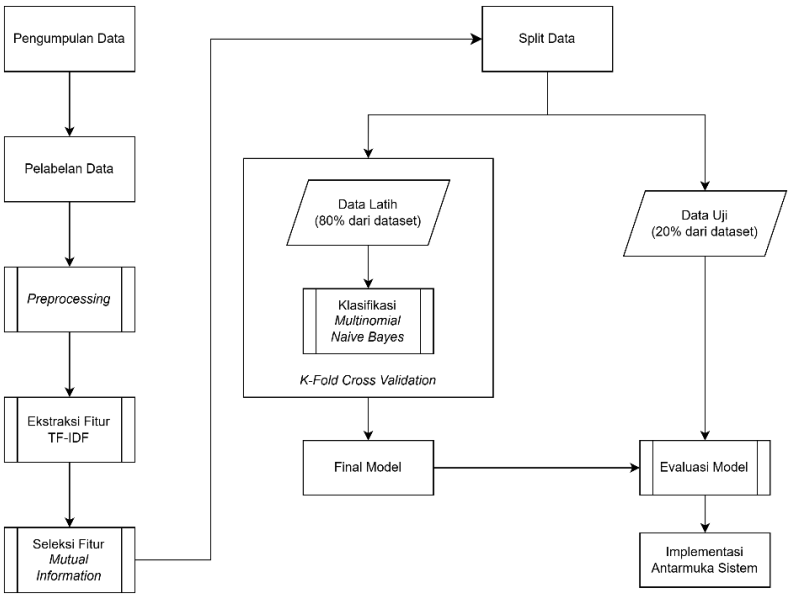
Permasalahan lainnya yaitu jika fitur atau kata yang digunakan terlalu banyak dan tidak memberikan informasi yang cukup signifikan sehingga menambah noise pada model dan menyebabkan prediksi

menjadi kurang akurat. Oleh sebab itu, proses seleksi fitur diperlukan untuk mengurangi fitur atau kata yang kurang informatif. Ernayanti dkk dalam penelitiannya membandingkan performa analisis sentimen dengan Multinomial Naive Bayes saat tidak menggunakan seleksi fitur dan performa saat menggunakan seleksi fitur. Algoritma Chi Square digunakan sebagai metode seleksi fitur pada penelitian ini dan hasilnya menunjukkan adanya peningkatan akurasi dari 88% tanpa menggunakan seleksi fitur menjadi 95% dengan seleksi fitur Chi Square [8]. Qodrin dkk dalam penelitiannya pun melakukan perbandingan serupa pada algoritma Support Vector Machine untuk mengevaluasi kinerja Mutual Information dalam proses seleksi fitur. Hasil penelitian menunjukkan bahwa penerapan Mutual Information memberikan performa yang lebih baik dengan akurasi sebelumnya yaitu 91% tanpa seleksi fitur menjadi 92% setelah menggunakan seleksi fitur [9].

Berdasarkan studi literatur yang telah diuraikan sebelumnya, diketahui bahwa baik penanganan negasi maupun seleksi fitur masing-masing dapat meningkatkan akurasi model analisis sentimen. Namun, penelitian yang menggabungkan kedua pendekatan tersebut masih jarang ditemukan. Hal ini yang mendorong penulis untuk menggabungkan penanganan negasi dan seleksi fitur Mutual Information serta mengevaluasi performanya. Penelitian ini dilakukan dengan menggunakan data ulasan atau review toko salah satu marketplace yaitu Shopee yang memiliki jumlah pengunjung terbanyak di Indonesia sepanjang tahun 2023**.** Penulis berharap dengan menggunakan kombinasi penanganan negasi dan seleksi fitur Mutual Information pada model Multinomial Naive Bayes dapat menghasilkan analisis sentimen yang lebih akurat.

2. Metode Penelitian

Penelitian dilakukan dalam beberapa tahap meliputi proses pengumpulan data, text preprocessing, ekstraksi fitur TF-IDF, seleksi fitur menggunakan Mutual Information, dan klasifikasi sentimen dengan algoritma Multinomial Naive Bayes. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Metode Penelitian

Dari alur penelitian yang ditunjukkan Gambar 1, proses penelitian secara umum dilakukan melalui tahapan pengumpulan data, pelabelan data, tahap *preprocessing*, ekstraksi fitur TF-IDF, seleksi fitur Mutual Information. Kemudian dilanjutkan dengan pembagian data menjadi data latih untuk melatih model dan data uji untuk mengevaluasi kinerja model.

2.1. Pengumpulan Data

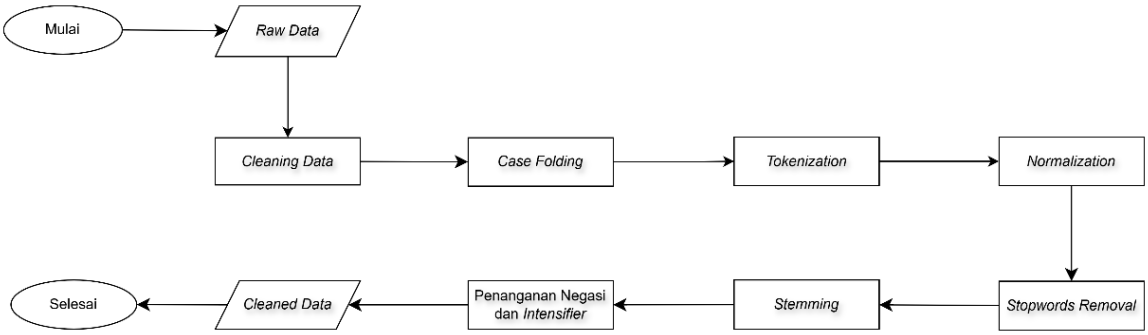
Data yang digunakan berasal dari ulasan berbagai produk baju atasan yang dijual oleh took *Ace Fashion Official Shop* di Shopee. Data diambil menggunakan metode *web scrapping* dengan Bahasa pemrograman Python dan hasilnya disimpan dalam bentuk file *excel*. Data yang diambil dalam proses

p-ISSN: 2301-5373
e-ISSN: 2654-5101

scrapping adalah *username*, jumlah bintang atau *rating*, dan ulasan yang diberikan oleh pengguna. *Rating* atau jumlah bintang akan digunakan sebagai acuan untuk melabeli data ulasan. Ulasan dengan *rating* satu dan dua akan dikategorikan ke dalam kelas sentimen negatif. Ulasan dengan *rating* tiga akan dikategorikan ke dalam kelas sentimen netral. Ulasan dengan *rating* empat dan lima akan dikategorikan ke dalam kelas sentimen positif. Adapun jumlah ulasan yang digunakan sebanyak 3000 data dengan pembagian 1000 data sentimen positif, 1000 data sentimen netral, dan 1000 data sentimen negatif.

2.2. *Text Preprocessing*

Text preprocessing adalah proses mengolah data teks mentah menjadi data yang lebih terstruktur sehingga data lebih siap untuk diproses pada tahap selanjutnya [10]. Tahapan dari *text preprocessing* ditunjukkan pada Gambar 2.



Gambar 2. Tahapan *Text Preprocessing*

Pada bagan yang dapat dilihat pada Gambar 2, *text preprocessing* diawali dengan proses *cleaning* data, yaitu menghapus angka, menghapus karakter non-alfanumerik, menghapus karakter yang bukan huruf, menghapus spasi yang berlebih, menghapus *newline* (\n), dan menghapus karakter *underscore*. Proses dilanjutkan pada tahap *case folding,* yaitu mengubah teks ke bentuk huruf kecil. Kemudian data melewati tahap *tokenization* membagi ulasan menjadi satuan token atau kata. Proses dilanjutkan ke tahap *normalization* yaitu proses memperbaiki token yang memiliki kesalahan eja atau *typo* serta membakukan kata yang tidak sesuai dengan kaidah Bahasa Indonesia. Selanjutnya yaitu proses *stopwords removal* yang menghapus kata-kata yang sering muncul tetapi tidak memberikan kontribusi signifikan terhadap akurasi dalam analisis sentimen. Selanjutnya yaitu *stemming* yang mengubah kata-kata berimbuhan menjadi bentuk kata dasar. Setelah itu, proses dilanjutkan pada tahap penanganan negasi dan *intensifier* yang bertujuan untuk menangani kata negasi atau kata *intensifier* yang dapat menyebabkan perubahan makna sentimen dari kalimat. Data hasil *text preprocessing* ditunjukkan pada Tabel 1.

Tabel 1. Contoh Hasil dari Tahapan *Text Preprocessing*

No	Tahapan	Hasil Pemrosesan
1	Data Mentah	Bagus bangett sesuai ekspetasi, ga rugi pokoknya belanja disini
2	*Case Folding* dan *Cleaning*	bagus bangett sesuai ekspetasi ga rugi pokoknya belanja disini
3	*Tokenization*	[bagus, bangett, sesuai, ekspetasi, ga, rugi, pokoknya, belanja, disini]
4	*Normalization*	[bagus, banget, sesuai, ekspektasi, tidak, rugi, pokoknya, belanja, sini]
5	*Stopwords Removal*	[bagus, banget, sesuai, ekspektasi, tidak, rugi, pokoknya, belanja, sini]
6	*Stemming*	[bagus, banget, sesuai, ekspektasi, tidak, rugi, pokok, belanja, sini]
7	Penangangan Negasi dan *Intensifier*	[bagus_banget, sesuai, ekspektasi, tidak, rugi, pokok, belanja, sini]

2.3. Ekstraksi Fitur TF-IDF

Ekstraksi fitur adalah tahapan yang mengkonversi *term* atau kata menjadi representasi angka yang dapat dimengerti oleh algoritma pemrosesan mesin. Salah satu metode ekstraksi fitur yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) yang bekerja dengan memberi nilai bobot setiap *term* dalam dokumen berdasarkan frekuensi kemunculan kata dalam dokumen [7]. Tahapan pembobotan menggunakan TF-IDF adalah sebagai berikut:

a. Menghitung nilai *Term Frequency* (TF)

$$tf_{t,d} = \frac{jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d}{total\ jumlah\ kata\ dalam\ dokumen\ d} \quad (1)$$

b. Menghitung nilai *Document Frequency* (DF)

$$df_t = jumlah\ dokumen\ yang\ mengandung\ kata\ t \quad (2)$$

c. Menghitung nilai *Inverse Document Frequency* (IDF)

$$idf_t = log_{10}\left(\frac{N}{df_t}\right) \quad (3)$$

d. Menghitung nilai TF-IDF

$$W_{t,d} = tf_{t,d} \times idf_t \quad (4)$$

Keterangan:

N = jumlah seluruh dokumen
$tf_{t,d}$ = nilai *TF* kata *t* dalam dokumen *d*
idf_t = nilai *IDF* kata *t*
$W_{t,d}$ = nilai bobot TF-IDF

2.4. Seleksi Fitur *Mutual Information*

Seleksi fitur merupakan tahapan untuk mempertahankan fitur atau kata yang relevan dan memberi informasi penting dalam membangun model klasifikasi serta mengabaikan fitur yang memiliki tingkat relevansi yang rendah dan bersifat *noise* sehingga proses membangun model menjadi lebih efektif karena hanya mempertahankan fitur yang penting [11]. *Mutual Information* melakukan seleksi fitur dengan menilai kontribusi setiap *term* terhadap keberadaan suatu kelas melalui perbandingan *score Mutual Information* dari *term* tersebut di setiap kelas [12]. Proses perhitungan *score Mutual Information* dapat dilakukan dengan menggunakan rumus berikut [13]:

$$I(U,C) = \frac{N_{11}}{N} log_2 \frac{N.N_{11}}{N_1.N_1} + \frac{N_{01}}{N} log_2 \frac{N.N_{01}}{N_0.N_1} + \frac{N_{10}}{N} log_2 \frac{N.N_{10}}{N_1.N_0} + \frac{N_{00}}{N} log_2 \frac{N.N_{00}}{N_0.N_0} \quad (5)$$

Keterangan:

N = total jumlah dokumen yang diamati untuk kombinasi term (*et*) dan kelas (*ec*), dimana N = N_{00} + N_{01} + N_{10} + N_{11}
N_{11} = jumlah dokumen yang mengandung term yang dicari (et) dan termasuk dalam kelas yang ditargetkan (ec)
N_{01} = jumlah dokumen yang tidak mengandung term yang dicari tetapi termasuk dalam kelas yang ditargetkan (ec)
N_{10} = jumlah dokumen yang mengandung term yang dicari namun berasal dari kelas lain (bukan kelas yang ditargetkan)
N_{00} = jumlah dokumen yang tidak mengandung term yang dicari dan tidak berasal dari kelas yang ditargetkan
$N_{1.}$ = total dokumen yang mengandung term yang dicari, yaitu hasil penjumlahan N_{10} dan N_{11}
N_1 = total dokumen yang termasuk dalam kelas yang ditargetkan, yaitu hasil penjumlahan N_{01} dan N_{11}
$N_{0.}$ = total dokumen yang tidak mengandung term yang dicari, yaitu hasil penjumlahan N_{01} dan N_{00}
N_0 = total dokumen yang tidak termasuk dalam kelas yang ditargetkan, yaitu hasil pennjumlahan N_{10} dan N_{00}

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.5. Klasifikasi *Multinomial Naive Bayes*

Algoritma *Multinomial Naive Bayes* bekerja dengan cara menghitung probabilitas fitur pada setiap kelas dan proses perhitungan untuk menentukan prediksi kelas pada data uji. Proses klasifikasi teks dengan algoritma *Multinomial Naive Bayes* terdiri dari beberapa tahap berikut [14]:

a. Menghitung probabilitas *prior*

$$P(c) = \frac{N(c)}{N} \quad (6)$$

Keterangan :

$P(c)$ = probabilitas atau peluang awal sebuah kelas sebelum mempertimbangkan fitur
$N(c)$ = jumlah dokumen yang muncul dalam kelas tertentu
N = total dokumen dalam dataset

b. Menghitung probabilitas kata unik

$$P(w|c) = \frac{count(w,c)+1}{count(c)+|V|} \quad (7)$$

$P(w|c)$ = probabilitas munculnya kata *w* dalam kelas *c*
$count(w,c)$ = jumlah kemunculan kata *w* dalam kelas *c*
$count(c)$ = jumlah total kata dalam kelas *c*
V = jumlah kata unik pada seluruh dokumen

c. Menghitung probabilitas dokumen atau kalimat

$$P\left(c\middle|d_{(n)}\right) = P(c) \times \prod P(w|c) \quad (8)$$

Keterangan :

$P\left(c\middle|d_{(n)}\right)$ = probabilitas atau peluang sebuah dokumen $d_{(n)}$ termasuk dalam kelas c
$P(c)$ = probabilitas awal dari kelas c
$\prod$ = operasi perkalian untuk seluruh kata yang muncul dalam dokumen
$P(w|c)$ = probabilitas kata w muncul pada kelas c

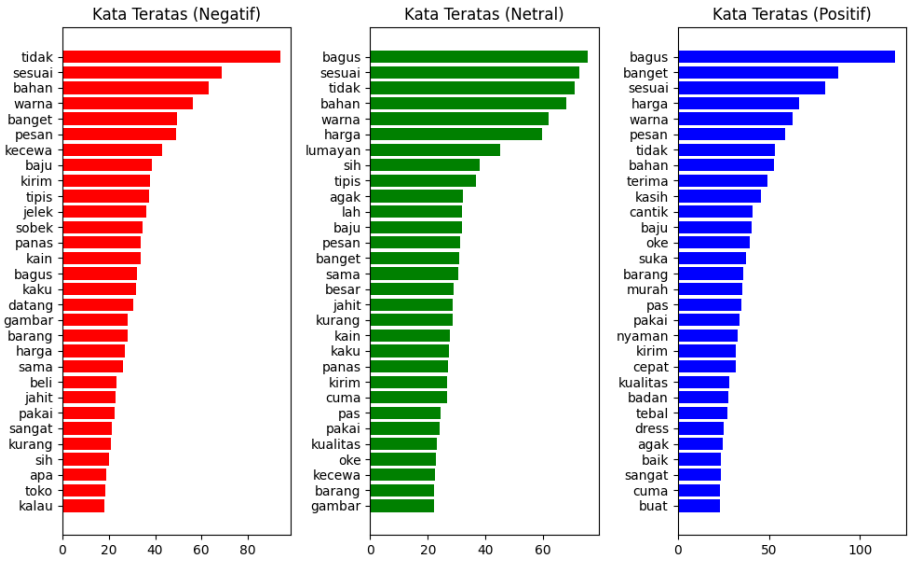
2.6. Skenario Pengujian dan Evaluasi

Terdapat empat skenario pengujian yang akan dilakukan. Skenario pertama, yaitu klasifikasi *Multinomial Naive Bayes* tanpa penanganan negasi dan *intensifier* dalam tahap *preprocessing* dan tanpa penggunaan seleksi fitur. Skenario pengujian pertama ini akan dijadikan sebagai metode *baseline*. Skenario kedua, yaitu klasifikasi dengan menerapkan penanganan negasi dan *intensifier* saja. Skenario ketiga, yaitu klasifikasi dengan menerapkan seleksi fitur saja. Skenario keempat, yaitu klasifikasi dengan menerapkan kombinasi penanganan negasi dan *intensifier* dengan penggunaan seleksi fitur. Empat skenario pengujian ini dirancang untuk mengkaji dampak penerapan penanganan negasi dan *intensifier* serta pengaruh penggunaan seleksi fitur pada analisis sentimen dengan algoritma *Multinomial Naive Bayes.* Sebelum pengujian dilakukan, data akan dibagi menjadi dua bagian yakni sebanyak 80% data menjadi data latih dan 20% dari data menjadi data uji. Evaluasi kinerja model dilakukan dengan metode *K-Fold Cross Validatio*n yang diterapkan hanya pada data latih dengan nilai *k* = 10. *K-Fold Cross Validation* membagi data latih menjadi 10 bagian (*fold*) yang sama besar. Pada setiap iterasi, satu fold dipilih sebagai data uji, sedangkan fold lainnya digunakan sebagai data latih. Proses ini diulang sebanyak *k* kali hingga setiap *fold* pernah menjadi data uji satu kali [15]. Performa model setiap skenario diukur dengan *confussion matrix* yang selanjutnya digunakan untuk mendapatkan nilai akurasi, *precision*, *recall,* dan *F1-Score*.

3. Hasil dan Pembahasan

3.1. Hasil Evaluasi Performa Metode *Baseline*

Skenario yang pertama yaitu pengujian model klasifikasi Multinomial Naive Bayes tanpa menggunakan penanganan negasi dan intensifier serta tidak melalui tahap seleksi fitur. Skenario ini dijadikan sebagai metode baseline untuk selanjutnya dibandingkan dengan hasil pengujian skenario lainnya. Namun sebelumnya, perlu diketahui distribusi kata-kata teratas untuk setiap kelas sentimen untuk mengetahui kata-kata yang dominan dan berperan penting saat proses klasifikasi sentimen. Visualisasi distribusi kata-kata teratas untuk setiap kelas sentimen berdasarkan nilai bobot TF-IDF ditunjukkan pada Gambar 3.



Gambar 3. Distribusi Kata Teratas Tiap Kelas pada Metode *Baseline*

Berdasarkan perbandingan distribusi kata yang mendominasi setiap kelas pada Gambar 3, diketahui bahwa terdapat beberapa kata yang mendominasi di kelas sentimen netral juga muncul dan mendominasi di kelas sentimen lainnya seperti kata “bagus”, “sesuai”, dan “tidak”. Apabila diperhatikan dengan seksama, kata “sesuai” dan kata “bagus” yang memiliki makna positif justru mendominasi kelas negatif. Kata yang sama mendominasi dua kelas sentimen serta kata yang memiliki makna berlawanan dengan kelas sentimen ini menunjukkan bahwa membedakan antara sentimen netral dengan sentimen positif atau sentimen negatif menjadi cukup sulit. Selain itu, terlihat kata yang tidak memiliki makna jika hanya terdiri dari satu kata saja juga cukup mendominasi seperti kata “banget” dan “sangat”. Evaluasi pada skenario pertama dilakukan dengan menghitung metrik seperti akurasi, precision, recall, dan F1-Score pada data validasi melalui proses *k-fold cross validation*. Hasil evaluasi skenario pertama ditunjukkan pada Tabel 2.

Tabel 2. Hasil Evaluasi Metode *Baseline*

Fold	Akurasi	*Precision*	*Recall*	*F1-Score*
1	0.7125	0.7114	0.7183	0.7133
2	0.6792	0.6851	0.6770	0.6777
3	0.6708	0.6710	0.6909	0.6740
4	0.6833	0.6900	0.6872	0.6870
5	0.7333	0.7380	0.7379	0.7379
6	0.7417	0.7379	0.7351	0.7356
7	0.7625	0.7592	0.7578	0.7581
8	0.7125	0.7282	0.7125	0.7166
9	0.7375	0.7470	0.7380	0.7378
10	0.7042	0.7061	0.6991	0.7002
Rata - rata	0.7137	0.7174	0.7154	0.7138

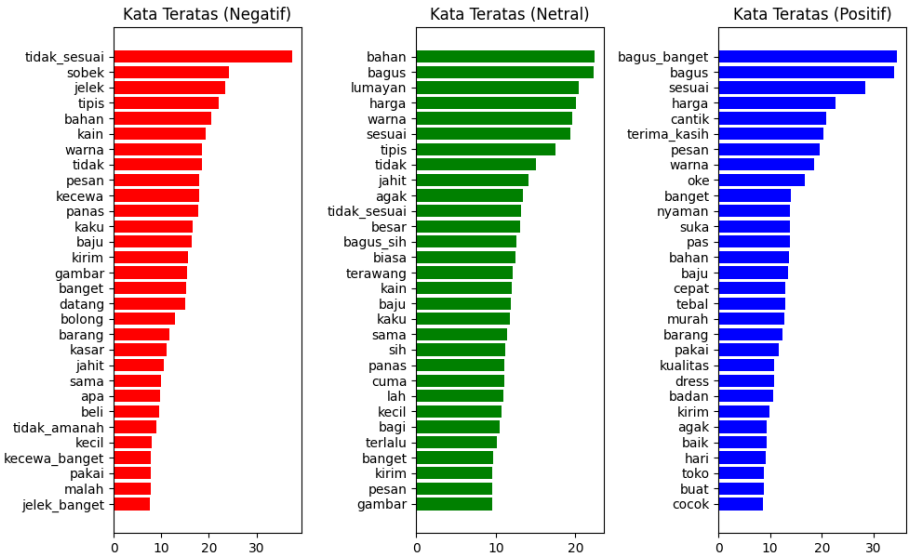
Berdasarkan hasil evaluasi pada Tabel 2, performa dari model Multinomial Naive Bayes pada skenario pertama yaitu tanpa penanganan negasi dan tanpa seleksi fitur menunjukkan performa yang cukup baik. Nilai rata-rata akurasi yaitu 0.7137 menunjukkan model cukup baik dalam memprediksi sentimen sesuai dengan label sentimen yang sebenarnya. Nilai precision dan recall berturut-turut yaitu sebesar

p-ISSN: 2301-5373
e-ISSN: 2654-5101

0.7174 dan 0.7154 menunjukkan model masih belum optimal dalam membedakan kelas dengan baik. Nilai F1-Score sebesar 0.7138 juga menunjukkan performa yang standar.

3.2. Hasil Evaluasi Performa Penanganan Negasi dan *Intensifier*

Skenario pengujian yang kedua yaitu pengujian model Multinomial Naive Bayes dengan penanganan negasi dan *intensifier*. Skenario ini bertujuan untuk mengetahui pengaruh penggunaan penanganan negasi dan intensifier dalam tahapan preprocessing terhadap akurasi model. Adapun visualisasi distribusi kata-kata teratas untuk setiap kelas sentimen yang melalui tahapan penanganan negasi dan *intensifier* ditunjukkan pada Gambar 4.



Gambar 4. Distribusi Kata Teratas Tiap Kelas dengan Penanganan Negasi dan *Intensifier*

Berdasarkan perbandingan distribusi kata yang mendominasi setiap kelas pada Gambar 4, terlihat adanya perubahan yang cukup signifikan pada kata-kata yang mendominasi setiap kelas setelah melalui tahap penanganan negasi dan *intensifier.* Distribusi kata di setiap kelas menjadi lebih spesifik dan berbeda satu sama lain. Pada kelas netral, muncul kata seperti “lumayan” dan “bagus_sih” yang dapat menjadi menjadi ciri khas kelas sentimen netral. Terdapat kata “sesuai” dan kata “tidak_sesuai” yang cukup imbang. Selain itu, pada kelas positif muncul kata “bagus_banget” dan “terima_kasih” yang menunjukkan ekspresi positif yang lebih kuat. Pada kelas negatif, muncul kata-kata seperti “tidak_sesuai”, “tidak_amanah”, “kecewa_banget” dan “jelek_banget” menunjukkan bahwa penanganan negasi mampu mengenali frasa bermakna negatif dengan lebih tepat. Penanganan negasi dan *intensifier* yang diterapkan mampu mengenali konteks kalimat dengan baik dan membuat perbedaan antar kelas sentimen terlihat lebih jelas. Pengujian skenario kedua dilakukan dengan langkah yang sama pada pengujian skenario pertama. Hasil evaluasi skenario pertama ditunjukkan pada Tabel 3.

Tabel 3. Hasil Evaluasi Penanganan Negasi dan *Intensifier*

Fold	Akurasi	*Precision*	*Recall*	*F1-Score*
1	0.7375	0.7353	0.7435	0.7372
2	0.6958	0.6948	0.6928	0.6924
3	0.6833	0.6842	0.7041	0.6884
4	0.7083	0.7188	0.7119	0.7128
5	0.7333	0.7425	0.7385	0.7393
6	0.7292	0.7237	0.7202	0.7212
7	0.7917	0.7859	0.7855	0.7850

Fold	Akurasi	*Precision*	*Recall*	*F1-Score*
8	0.7417	0.7514	0.7321	0.7434
9	0.7417	0.7530	0.7418	0.7425
10	0.7500	0.7455	0.7422	0.7426
Rata - rata	0.7312	0.7335	0.7323	0.7305

Berdasarkan hasil evaluasi *K-Fold Cross Validation* pada Tabel 3, performa dari model Multinomial Naive Bayes pada skenario kedua yaitu model Multinomial Naive Bayes dengan penanganan negasi dan intensifier serta tanpa seleksi fitur menunjukkan adanya peningkatan performa dibandingkan dengan skenario pengujian pertama. Rata-rata akurasi meningkat dari 0.7137 menjadi 0.7312. Nilai precision mengalami peningkatan dari yang sebelumnya sebesar 0.7174 menjadi 0.7335. Nilai recall dan F1-Score juga mengalami peningkatan yang awalnya sebesar 0.7154 dan 0.7138 menjadi 0.7323 dan 0.7305. Peningkatan nilai akurasi, precision, recall, dan F1-Score ini menunjukkan bahwa penanganan negasi dan intensifier mampu membantu model dalam memahami konteks makna kata atau kalimat ulasan dengan lebih tepat.

3.3. Hasil Evaluasi Performa Seleksi Fitur

Skenario pengujian yang ketiga yaitu pengujian model Multinomial Naive Bayes dengan tahapan seleksi fitur Mutual Information tanpa penanganan negasi dan intensifier. Skenario ini bertujuan untuk mengetahui pengaruh penggunaan seleksi fitur terhadap akurasi model. Pada pengujian ini, persentase jumlah fitur dengan bobot atau nilai score Mutual Information teratas yang dipertahankan berada pada rentang 10% hingga 90%. Hasil evaluasi skenario ketiga ditunjukkan pada Tabel 4.

Tabel 4. Hasil Evaluasi Seleksi Fitur

Jumlah Fitur (%)	Rata - Rata			
	Akurasi	*Precision*	*Recall*	*F1-Score*
10	0.7071	0.7103	0.7096	0.7057
20	0.7167	0.7203	0.7186	0.7166
30	0.7183	0.7214	0.7200	0.7180
40	0.7163	0.7189	0.7178	0.7158
50	0.7146	0.7165	0.7163	0.7139
60	0.7163	0.7177	0.7181	0.7155
70	0.7167	0.7184	0.7187	0.7160
80	0.7158	0.7176	0.7177	0.7151
90	0.7154	0.7181	0.7171	0.7151

Berdasarkan hasil evaluasi pada Tabel 4, diketahui bahwa performa terbaik model diperoleh saat menggunakan seleksi fitur yang mempertahankan 30% fitur dari total seluruh fitur dengan akurasi 0.7183, *precision* 0.7214, *recall* 0.7200, dan *F1-Score* 0.7180. Hasil ini membuktikan bahwa proses seleksi fitur mampu menyaring kata-kata yang paling relevan dalam klasifikasi dan membantu model dalam mengidentifikasi pola sentimen tanpa harus memproses seluruh fitur dalam data. Namun, ketika jumlah fitur yang dipertahankan dalam proses klasifikasi semakin banyak atau menggunakan hampir seluruh fitur, performa model cenderung menurun. Hal ini menunjukkan bahwa terlalu banyak fitur yang kurang informatif dan bersifat *noise* dapat menurunkan kinerja model dalam proses klasifikasi. Penggunaan fitur yang terlalu sedikit juga menyebabkan kinerja model menjadi kurang maksimal karena terbatasnya informasi yang diberikan ke model.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.4. Hasil Evaluasi Performa Kombinasi Penanganan Negasi dan *Intensifier* dengan Seleksi Fitur

Skenario pengujian yang keempat yaitu pengujian model *Multinomial Naive Bayes* dengan kombinasi dari penggunaan penanganan negasi dan *intensifier* serta penggunaan seleksi fitur. Skenario ini bertujuan untuk mengetahui pengaruh penggunaan penanganan negasi dan *intensifier* pada tahapan *preprocessing* serta penggunaan seleksi fitur terhadap akurasi model. Pengujian skenario keempat ini dilakukan dengan menghitung metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-Score* dengan berbagai persentase jumlah fitur yang dipertahankan dari 10% hingga 90% atau menggunakan seluruh fitur. Hasil evaluasi skenario keempat ditunjukkan pada Tabel 5.

Tabel 5. Hasil Evaluasi Penanganan Negasi dan *Intensifier* dengan Seleksi Fitur

Jumlah Fitur (%)	Rata - Rata			
	Akurasi	*Precision*	*Recall*	*F1-Score*
10	0.7171	0.7186	0.7186	0.7143
20	0.7375	0.7402	0.7388	0.7367
30	0.7371	0.7388	0.7381	0.7361
40	0.7383	0.7417	0.7391	0.7377
50	0.7400	0.7431	0.7412	0.7395
60	0.7408	0.7432	0.7419	0.7404
70	0.7387	0.7405	0.7397	0.7381
80	0.7396	0.7419	0.7407	0.7390
90	0.7379	0.7410	0.7391	0.7374

Berdasarkan hasil akurasi pada Tabel 5, nilai akurasi tertinggi didapatkan saat model mempertahankan 60% fitur dari keseluruhan fitur dengan nilai akurasi sebesar 0.7408, *precision* sebesar 0.7432, *recall* sebesar 0.7419, dan *F1-Score* sebesar 0.7404. Jika dibandingkan dengan hasil evaluasi skenario 2 dengan nilai akurasi 0.7312, *precision* 0.7335, *recall* 0.7323, dan *F1-Score* 0.7305 serta hasil evaluasi skenario 3 dengan nilai akurasi 0.7183, *precision* 0.7214, *recall* 0.7200, dan *F1-Score* 0.7180, maka hasil evaluasi pada skenario 4 ini memberikan peningkatan akurasi yang signifikan. Hal ini menunjukkan kombinasi penanganan negasi dan *intensifier* bersama seleksi fitur *Mutual Information* memberikan performa yang lebih baik daripada hanya menerapkan penanganan negasi dan *intensifier* saja maupun hanya menerapkan penggunaan seleksi fitur saja. Skenario 4 menunjukkan bahwa penanganan negasi dan *intensifier* berhasil mengenali konteks kalimat dengan lebih baik sehingga perbedaan ciri khas antar kelas sentimen menjadi lebih jelas. Sementara itu, seleksi fitur *Mutual Information* mampu menyaring kata-kata yang paling relevan dalam klasifikasi dan membantu model dalam mengidentifikasi pola sentimen tanpa harus memproses seluruh fitur dalam data.

3.5. Performa Model Terbaik

Hasil evaluasi dari setiap skenario pengujian kemudian dibandingkan untuk mengetahui model dengan akurasi atau performa terbaik. Skenario 1 merupakan metode baseline yang tidak menggunakan penanganan negasi dan intensifier maupun seleksi fitur. Skenario 2 merupakan model dengan penanganan negasi dan intensifier saja. Skenario 3 merupakan model dengan penggunaan seleksi fitur saja. Skenario 4 merupakan model dengan kombinasi keduanya. Adapun hasil evaluasi terbaik dari setiap skenario pengujian dapat dilihat pada Tabel 6.

Tabel 6. Perbandingan Evaluasi dari Setiap Skenario Pengujian

Skenario Pengujian	Akurasi	*Precision*	*Recall*	*F1-Score*
1	0.7137	0.7174	0.7154	0.7138
2	0.7312	0.7335	0.7323	0.7305

Skenario Pengujian	Akurasi	*Precision*	*Recall*	*F1-Score*
3	0.7183	0.7214	0.7200	0.7180
4	0.7408	0.7432	0.7419	0.7404

Berdasarkan perbandingan hasil evaluasi pada Tabel 6, model dengan performa terbaik ditunjukkan oleh model pada skenario pengujian keempat yaitu model *Multinomial Naive Bayes* yang menggunakan kombinasi penanganan negasi dan intensifier serta seleksi fitur *Mutual Information* dengan jumlah fitur yang dipertahankan sebanyak 60% dari total fitur. Penanganan negasi dan *intensifier* dapat membantu model memahami konteks kata atau kalimat dengan lebih tepat terutama akibat adanya kata negasi yang membalikkan makna kata dan kata *intensifier* yang memperkuat sentimen. Sementara seleksi fitur *Mutual Information* bertugas untuk memilih kata-kata yang paling relevan dan berpengaruh dalam klasifikasi tanpa harus memproses seluruh fitur yang terkadang kurang informatif dan bersifat *noise*. Nilai akurasi yang didapatkan pada skenario pengujian keempat ini menunjukkan bahwa model dengan kombinasi penanganan negasi dan *intensifier* dengan penggunaan seleksi fitur mampu meningkatkan kinerja model menjadi lebih baik.

4. Kesimpulan

Berdasarkan hasil penelitian, diketahui bahwa penanganan negasi dan intensifier pada tahapan preprocessing data mampu meningkatkan akurasi dari 71,37% menjadi 73,12%, precision dari 71,74% menjadi 73,35%, recall dari 71,54% menjadi 73,23%, dan F1-Score dari 71,38% menjadi 73,05%. Hal ini menunjukkan bahwa penanganan negasi dan intensifier mampu membantu model memahami konteks kata atau kalimat dengan lebih tepat terutama akibat adanya kata negasi yang membalikkan makna kata dan kata intensifier yang dapat memperkuat atau memperlemah sentimen.

Adapun penerapan seleksi fitur Mutual Information mampu menyaring fitur atau kata yang paling relevan dalam analisis sentimen dan mengabaikan fitur yang kurang informatif sehingga model dapat mengidentifikasi pola sentimen dengan efektif tanpa harus memproses seluruh fitur atau kata dalam data. Hal ini ditunjukkan dengan adanya peningkatan akurasi dari 71,37% menjadi 71,83%, precision dari 71,74% menjadi 72,14%, recall dari 71,54% menjadi 72,00%, dan F1-Score dari 71,38% menjadi 71,80% saat mempertahankan fitur sebanyak 30% dari seluruh fitur.

Performa model terbaik didapatkan saat mengkombinasikan penanganan negasi dan intensifier dengan seleksi fitur Mutual Information dengan peningkatan akurasi dari 71,37% menjadi 74,08%, precision dari 71,74% menjadi 74,32%, recall dari 71,54% menjadi 74,19%, dan F1-Score dari 71,38% menjadi 74,04% saat mempertahankan jumlah fitur sebesar 60% dari total data. Hal ini menunjukkan bahwa kombinasi penanganan negasi dan *intensifier* bersama seleksi fitur *Mutual Information* memberikan performa yang lebih baik daripada hanya menerapkan penanganan negasi dan *intensifier* saja maupun hanya menerapkan penggunaan seleksi fitur saja.

Daftar Pustaka

[1] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*. 2023.
[2] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, “Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square,” *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
[3] U. Muhammadiyah Jember, M. Izunnahdi, G. Aburrahman, and A. Eko Wardoyo, “Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google PlayStore Menggunakan Metode Multinomial Naive Bayes Sentiment Analysis on KAI Access Application Review Data on Google PlayStore Using Multinomial Naive Bayes Method,” *J. Smart Teknol.*, vol. 4, no. 2, pp. 2774–1702, 2023, [Online]. Available: http://jurnal.unmuhjember.ac.id/index.php/JST
[4] S. Mandasari, B. H. Hayadi, and R. Gunawan, “Analisis Sentimen Pengguna Transportasi Online Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 2, p. 118, 2022, doi: 10.53513/jsk.v5i2.5635.
[5] R. I. Tarecha, F. Wahyudi, and U. M. Jannah, “Penanganan Negasi dalam Analisa Sentimen Bahasa Indonesia,” *J. Sist. Inf. dan Inform.*, vol. 1, no. 1, pp. 51–58, 2022, doi: 10.33379/jusifor.v1i1.1276.
[6] N. Ananda, M. Fikry, L. Handayani, and I. Iskandar, “Klasifikasi Sentimen Tweet Masyarakat

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Terhadap Kendaraan Listrik Menggunakan Support Vector Machine," vol. 8, no. 4, pp. 568–577, 2023, doi: 10.32493/informatika.v8i4.36754.

[7] N. I. Darmawan and B. D. Setiawan, "Analisis Sentimen terhadap Akuisisi Saham Tokopedia oleh TikTok Menggunakan Naive Bayes berdasarkan Komentar YouTube," vol. 1, no. 1, pp. 1–8, 2017.

[8] T. Ernayanti, M. Mustafid, A. Rusgiyono, and A. R. Hakim, "Penggunaan Seleksi Fitur Chi-Square Dan Algoritma Multinomial Naive Bayes Untuk Analisis Sentimen Pelangggan Tokopedia," *J. Gaussian*, vol. 11, no. 4, pp. 562–571, 2023, doi: 10.14710/j.gauss.11.4.562-571.

[9] S. Al Qodrin, N. Yusliani, and A. Syahrini, "Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine dan Seleksi Fitur Mutual Information," *J. JUPITER*, vol. 14, no. 2, p. 44, 2022.

[10] S. Shevira, I. M. A. D. Suarjaya, and P. W. Buana, "Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia," *JITTER J. Ilm. Teknol. dan Komput.*, vol. 3, no. 2, p. 1074, 2022, doi: 10.24843/jtrti.2022.v03.i02.p06.

[11] D. B. W. Alfredo Gormantara, "Klasifikasi Kategori dan Pelabelan Berita Bahasa Indonesia Menggunakan Mutual Information Dan K-Nearest Neighbor," *Tematika2*, vol. 8, pp. 75–82, 2020.

[12] N. Y. Abel Filemon Haganta Kaban, Indriati, "Analisis Sentimen Aplikasi E-Goverment berdasarkan Ulasan Pengguna menggunakan Metode Maximum Entropy dan Seleksi Fitur Mutual Information," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 1452–1458, 2021.

[13] I. G. Bgs, D. Putra, and C. Pramartha, "Penerapan SVM dengan Seleksi Fitur Mutual Information untuk Memprediksi Penerapan SVM dengan Seleksi Fitur Mutual Information untuk Memprediksi Sentimen PEMILU 2024," no. May, 2024.

[14] F. E. Purwiantono and A. Aditya, "Klasifikasi Sentimen Sara, Hoaks Dan Radikal Pada Postingan Media Sosial Menggunakan Algoritma Naive Bayes Multinomial Text," *J. Tekno Kompak*, vol. 14, no. 2, p. 68, 2020, doi: 10.33365/jtk.v14i2.709.

[15] R. N. Irawan, K. M. Hindrayani, and M. Idhom, "Penerapan Cross Validation sebagai Analisis Sentimen Pelayanan Publik Kereta Api Lokal Daop 8 Menggunakan Metode Multinomial Naive Bayes," *G-Tech J. Teknol. Terap.*, vol. 8, no. 2, pp. 954–963, 2024, doi: 10.33379/gtech.v8i2.4117.

This page is intentionally left blank.