

Analisis Perbandingan Naïve Bayes dan Random Forest untuk Evaluasi PKKMB Berbasis Sentimen Mahasiswa

Comparative Analysis of Naïve Bayes and Random Forest for Student Sentiment-Based PKKMB Evaluation

Alexander Sole*¹, Anis Zubair²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Merdeka Malang

*e-mail: solesandro981@gmail.com¹, anis.zubair@unmer.ac.id²

Received: 11.12.2025	Revised: 08.02.2026	Accepted: 09.03.2026	Available online: 05.04.2026
-------------------------	------------------------	-------------------------	---------------------------------

Abstrak: Pengenalan Kehidupan Kampus Mahasiswa Baru (PKKMB) adalah kegiatan penting yang membantu mahasiswa baru mengenal lingkungan akademik dan budaya kampus. Evaluasi terhadap pelaksanaan PKKMB umumnya dilakukan melalui survei numerik yang hanya memberikan gambaran umum sehingga kurang mampu menangkap opini mahasiswa secara mendalam. Penelitian ini menerapkan analisis sentimen berbasis machine learning untuk mengolah 886 data komentar evaluasi PKKMB di Universitas Merdeka Malang. Proses penelitian meliputi tahapan preprocessing teks (case folding, cleansing, tokenization, normalisasi, stopword removal, dan stemming), pelabelan otomatis menggunakan model IndoBERT, serta konversi teks menjadi fitur numerik menggunakan TF-IDF. Dua algoritma klasifikasi, yaitu Naïve Bayes dan Random Forest, digunakan untuk membandingkan performa model. Data dibagi menggunakan rasio 80:20 untuk data latih dan data uji, serta evaluasi model dilakukan menggunakan matriks accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Naïve Bayes memiliki performa yang lebih unggul dibandingkan Random Forest dengan nilai accuracy 81,4%, precision 83,04%, recall 86,92%, dan F1-score 84,93%. Sementara itu, Random Forest menghasilkan accuracy 79,78%, precision 83,18%, recall 83,18%, dan F1-score 83,18%. Temuan ini menunjukkan bahwa Naïve Bayes merupakan metode yang lebih efektif untuk mengklasifikasikan sentimen positif dan negatif pada dataset evaluasi PKKMB dengan pelabelan otomatis. Hasil analisis ini juga memberikan gambaran aspek-aspek kegiatan PKKMB yang masih perlu diperbaiki sebagai rekomendasi peningkatan pelaksanaan di tahun berikutnya.

Kata kunci: Analisis Sentimen, Naïve Bayes, Random Forest, PKKMB, TF-IDF.

Abstract: Introduction to New Student Campus Life (PKKMB) program is an essential activity designed to help new students become familiar with the academic environment and campus culture. Evaluations of PKKMB are generally conducted using numerical surveys, which provide only a general overview and are often insufficient to capture students' opinions in depth. This study applies machine learning-based sentiment analysis to process 886 textual comments from PKKMB evaluation data at Merdeka University Malang. The research process includes several text preprocessing stages (case folding, cleansing, tokenization, normalization, stopword removal, and stemming), automatic labeling using the IndoBERT model, and text-to-feature transformation using TF-IDF. Two classification algorithms, Naïve Bayes and Random Forest, are employed to compare model performance. The dataset is divided using an 80:20 ratio for training and testing, and model evaluation is conducted using accuracy, precision, recall, and F1-score metrics. The results show that Naïve Bayes outperforms Random Forest with an accuracy of 81.4%, precision of 83.04%, recall of 86.92%, and F1-score of 84.93%. Meanwhile, Random Forest achieves an accuracy of 79.78%, precision of 83.18%, recall of 83.18%, and an F1-score of 83.18%. These findings indicate that Naïve Bayes is a more effective method for classifying positive and negative sentiments in the PKKMB evaluation dataset using automated labeling. The sentiment analysis results also highlight aspects of the PKKMB program that require improvement, serving as recommendations for enhancing its implementation in future years.

Keywords: Sentiment Analysis, Naïve Bayes, Random Forest, PKKMB, TF-IDF

1. PENDAHULUAN

Pengenalan Kehidupan Kampus Mahasiswa Baru (PKKMB) merupakan kegiatan resmi tahunan yang bertujuan untuk memperkenalkan mahasiswa baru pada lingkungan akademik, sistem pendidikan serta peraturan yang berlaku di perguruan tinggi [1]. Universitas Merdeka Malang Merupakan salah satu Perguruan tinggi swasta di Jawa Timur yang menerapkan PKKMB. PKKMB menjadi tahap awal dalam membentuk persepsi mahasiswa baru terhadap institusi dan membantu proses adaptasi di lingkungan perguruan tinggi. Namun, persepsi mahasiswa terhadap pelaksanaan PKKMB dapat bervariasi dimana sebagian mahasiswa menilai kegiatan ini bermanfaat, sementara

sebagian lain menganggap masih terdapat kekurangan dalam pelaksanaannya. Untuk mengevaluasi kegiatan ini, perguruan tinggi umumnya menggunakan survey atau kuesioner yang menghasilkan data kuantitatif, seperti nilai rata-rata kepuasan atau persentase tanggapan. Meskipun memberikan gambaran umum, pendekatan ini cenderung bersifat deskriptif sederhana dan kurang mampu menangkap opini serta sentimen mahasiswa secara mendalam. Padahal, komentar teks pada survey evaluasi mengandung informasi penting berupa opini, kritik atau saran yang dapat dimanfaatkan dalam meningkatkan kualitas pelaksanaan PKKMB.

Untuk mengatasi keterbatasan tersebut, analisis sentimen menjadi solusi dalam pengolahan data berupa teks untuk memahami emosi atau pendapat yang disampaikan [2]. Analisis sentimen merupakan bagian dari teks mining yang bertujuan untuk mengetahui kecenderungan sentimen positif atau negatif secara objektif. Proses ini memanfaatkan teknik Natural Language Processing (NLP) dalam mengolah bahasa alami melalui tahapan pemrosesan teks hingga dapat dianalisis polanya [3], [4]. Analisis sentimen telah banyak diterapkan di berbagai bidang, seperti ulasan produk dan media sosial.

Pada ulasan produk menunjukkan metode Random Forest memiliki performa lebih baik dibanding Naïve Bayes pada analisis sentimen ulasan aplikasi BCA Mobile dengan akurasi 93,93%, recall 89,89% dan F1-score 91,43% [5]. Sementara itu, dalam konteks media sosial, Miftahusalam et al. menganalisis sentimen twitter tentang penghapusan tenaga honorer menggunakan Random Forest, Naïve Bayes dan Support Vector Machine. Hasilnya menunjukkan Random Forest menjadi metode terbaik dengan akurasi random oversampling sebesar 66% dan F1-Score sebesar 67 % [6]. Penelitian lain juga membandingkan metode Naïve Bayes, Decision Tree, Random Forest pada sentimen biaya haji media sosial Youtube tahun 2023, hasilnya metode Naïve Bayes lebih unggul dari metode Decision Tree dan Random Forest dengan akurasi 90% [7].

Seperti ulasan produk dan media sosial, evaluasi PKKMB juga menghasilkan data teks berupa komentar mahasiswa yang kaya akan opini. Analisis sentimen dengan pendekatan machine learning dapat digunakan untuk mengklasifikasi sentimen tersebut secara lebih mendalam. Dua algoritma yang umum digunakan adalah Naïve Bayes dan Random Forest. Naïve Bayes ialah cara pengklasifikasian berbasis probabilitas yang sederhana serta cepat dan efisien dalam pengolahan teks [8]. Sebaliknya, Random Forest sebagai algoritma berbasis ensemble yang mengkombinasikan banyak pohon keputusan untuk mengelola data yang kompleks, sehingga sering kali menghasilkan akurasi yang lebih tinggi [9]. meskipun kedua algoritma ini memiliki keunggulan masing-masing, belum ada penelitian di universitas merdeka malang yang secara khusus membandingkan kinerja keduanya dalam analisis sentimen terhadap data evaluasi PKKMB.

Ketiadaan penelitian tersebut, menunjukkan adanya kesenjangan dalam evaluasi PKKMB yang selama ini sering mengandalkan pendekatan deskriptif sederhana. Maka dari itu, tujuan penelitian ini untuk membandingkan kinerja algoritma Naïve Bayes dan Random Forest untuk menentukan metode terbaik serta memberikan rekomendasi evaluasi dan perbaikan berdasarkan analisis sentimen data evaluasi PKKMB di Universitas Merdeka Malang.

2. TINJAUAN PUSTAKA

2.1. *Natural Language Processing (NLP)*

Natural Language Processing (NLP) merupakan cabang ilmu komputer dan kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa alami. NLP memungkinkan komputer untuk memahami, memproses dan menghasilkan teks yang mirip dengan komunikasi manusia [3]. Proses NLP terdiri dari beberapa tahapan yakni tokenisasi, case folding, stopword removal, stemming dan representasi teks atau mengubah format teks menjadi numerik proses ini dikenal sebagai text processing. Hasil dari tahapan NLP ini kemudian digunakan oleh teks mining untuk menganalisis data.

2.2. *Text Mining*

Teks mining merupakan proses menemukan pola, hubungan atau informasi penting dari sekumpulan data teks yang tidak terstruktur. Menurut Putra dan Wardani [10] Text mining merupakan proses otomatis untuk menggali informasi dari dokumen teks serta mengidentifikasi

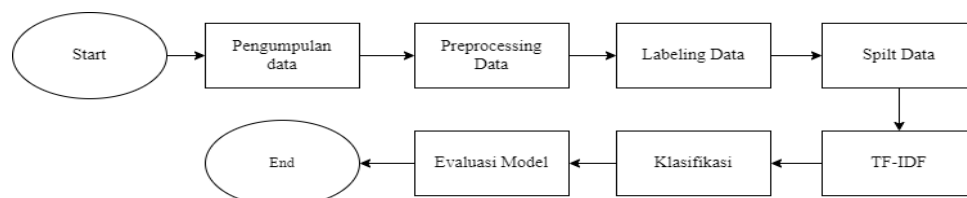
hubungan antar dokumen menggunakan metode statistik seperti association, clustering dan classification.

2.3. Analisis Sentimen

Sentimen atau ekspresi merupakan bentuk perasaan, opini atau sikap emosional seseorang terhadap suatu topik tertentu [11]. Analisis sentimen atau opinion mining merupakan sub bagian dari text mining yang memanfaatkan teknik Natural Language Processing (NLP) untuk mengolah bahasa alami. Tujuan utama dari analisis sentimen adalah mengidentifikasi, mengklasifikasi emosi atau opini dalam teks, umumnya dikategorikan kedalam sentimen positif dan negatif [4]. Melalui proses ini sistem dapat memberikan gambaran, pandangan atau sikap seseorang terhadap topik tertentu secara objektif dan terukur.

3. METODE

Metodologi penelitian ini disusun dalam serangkaian langkah-langkah, dimulai dengan Proses Pengumpulan data sampai pada evaluasi model seperti pada Gambar 1. Setiap tahapan diuraikan sebagai berikut:



Gambar 1. Alur Penelitian

3.1 Pengumpulan Data

Data yang dipakai pada penelitian ini adalah data sekunder yang didapat dari Biro Kemahasiswaan dan Alumni (BKA) Universitas Merdeka Malang berupa hasil survei evaluasi PKKMB tahun akademik 2025. Dataset tersebut memiliki beberapa atribut, namun penelitian ini hanya menggunakan atribut berupa opini, kritik maupun saran terhadap pelaksanaan PKKMB. Data awal terdiri dari 926 komentar teks mahasiswa baru, setelah dilakukan proses validasi dan pembersihan data kosong diperoleh 886 data layak pakai.

3.2 Preprocessing Data

Preprocessing data merupakan tahapan awal untuk membersihkan dan menyeragamkan teks agar sesuai dengan kebutuhan analisis. Tahapan ini penting dalam analisis sentimen karena bertujuan menghilangkan *noise*, mengurangi dimensi data serta membuat data lebih terstruktur sehingga dapat diproses oleh algoritma *machine learning* [12]. Dalam *preprocessing* data terdapat beberapa proses didalamnya, diantaranya yakni *case folding*, *tokenization*, *cleansing*, *normalization*, *stopword removal* dan *stemming* [13].

a. Case folding

Case folding adalah langkah-langkah penyamaan bentuk huruf dengan cara merubah semua karakter menjadi huruf kecil (*lowercase*).

b. Tokenization

Tokenization adalah tahapan dalam preprocessing untuk memecah teks menjadi potongan-potongan kecil yang dinamakan token. Mekanisme ini dilakukan untuk memisahkan teks menjadi elemen-elemen yang dapat diproses secara individual.

c. Cleansing

Cleansing ialah penghilangan elemen yang tidak sesuai dari teks seperti angka, *emoticon*, tanda baca, simbol khusus maupun karakter yang bukan alfabet lainnya.

d. Normalization

Normalization yaitu alur penyeragaman terhadap kata-kata yang nonformal seperti penggunaan bahasa daerah, bahasa gaul maupun singkatan dengan cara mengubah kata tersebut menjadi kata baku berdasarkan kaidah KBBI.

e. Stopword removal

Stopword removal ialah penghilangan kata-kata fungsional namun tidak memiliki makna yang penting seperti kata hubung, kata depan atau preposisi dengan tujuan untuk memangkas jumlah teks yang akan diolah sehingga mekanisme yang dilakukan lebih cepat dan tepat sasaran.

f. *Stemming*

Stemming ialah bagian pra pemrosesan yang memiliki tujuan untuk mengubah kata ke dalam bentuk dasarnya dengan menghapus atau menghapus imbuhan (awalan, akhiran, dan sisipan) pada kata.

3.3 Labeling Data

Proses labeling adalah tahapan pemberian kategori di setiap data guna mengidentifikasi sentimen positif maupun negatif. Proses ini dilaksanakan secara otomatis dengan bahasa pemrograman python dengan memakai model *IndoBERT*. Model yang digunakan merupakan model publik yang diperoleh dari *Hugging Face Hub* dengan nama *agussamudra/indo-sentiment-analysis* yang diimplementasikan melalui fitur pipeline pada *library transformers*. Label yang dihasilkan berupa LABEL_0 untuk sentimen negatif dan LABEL_1 untuk sentimen positif.

3.4 Split Data

Split data ialah teknik pada *machine learning* untuk memisahkan dataset menjadi dua subset yaitu data latih (*training data*) dan data uji (*testing data*) [14]. Training data dipakai untuk membangun model, sedangkan testing data diterapkan untuk mengetes pengklasifikasian yang sebelumnya dibuat menggunakan data training [15]. Pada penelitian ini, penyaluran data dilaksanakan secara acak (*random sampling*) dengan rasio 80:20, dimana 80% data digunakan sebagai data latih dan 20% digunakan sebagai data uji. Membagi secara acak bertujuan untuk menjaga keseimbangan distribusi kelas sentimen sehingga model tidak bias terhadap kelas tertentu.

3.5 TF-IDF

TF-IDF yaitu model representasi teks yang dipergunakan untuk mentransformasi teks menjadi fitur numerik. Metode ini mengkombinasikan dua pendekatan pembobotan kata, yaitu *term frequency (TF)* dan *inverse document frequency (IDF)*. TF mengukur seberapa sering suatu kata muncul dalam dokumen tertentu. TF dihitung sebagai rasio jumlah timbulnya kata terhadap total kata dalam dokumen [16]. IDF Mengukur frekuensi kelangkaan sebuah kata muncul di seluruh kumpulan dokumen. Dimana kata yang jarang muncul diberikan bobot lebih besar karena dinilai memiliki informasi yang lebih penting [17]. Proses ini dapat diterapkan dengan persamaan berikut:

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{Total kata dalam dokumen}} \quad (1)$$

$$IDF = \log \frac{\text{jumlah total dokumen}}{\text{jumlah dokumen yang mengandung kata}} + 1 \quad (2)$$

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

Dalam penelitian ini, untuk memberikan contoh cara kerja perhitungan TF-IDF, peneliti menggunakan tiga dokumen contoh data, yang dapat dilihat pada Tabel 1.

Tabel 1. Contoh Ulasan

Dokumen	Teks sebelum Pembersihan	Teks setelah pembersihan
D1	saran saya pkkmb kedepannya lebih seru lagi	saran pkkmb depan lebih seru
D2	biar lebih seru dan tidak membosankan perbanyak goyang goyangnya	biar lebih seru tidak bosan banyak goyang goyang
D3	pkkmb kedepannya di kasih konsumsi	pkkmb depan kasih konsumsi

Data yang telah melewati proses pembersihan, selanjutnya dihitung frekuensi kemunculan kata dalam dokumen melalui proses TF pada Tabel 2.

Tabel 2. Perhitungan TF

kata(term)	Kemunculan Kata			TF		
	D1	D2	D3	D1	D2	D3
saran	1	0	0	$1/5 = 0,200$	$0/8 = 0$	$0/4 = 0$
pkkmb	1	0	1	$1/5 = 0,200$	$0/8 = 0$	$1/4 = 0,250$
depan	1	0	1	$1/5 = 0,200$	$0/8 = 0$	$1/4 = 0,250$
lebih	1	1	0	$1/5 = 0,200$	$1/8 = 0.125$	$0/4 = 0$
seru	1	1	0	$1/5 = 0,200$	$1/8 = 0.125$	$0/4 = 0$
biar	0	1	0	$0/5 = 0$	$1/8 = 0.125$	$0/4 = 0$
tidak	0	1	0	$0/5 = 0$	$1/8 = 0.125$	$0/4 = 0$
bosan	0	1	0	$0/5 = 0$	$1/8 = 0.125$	$0/4 = 0$
banyak	0	1	0	$0/5 = 0$	$1/8 = 0.125$	$0/4 = 0$
goyang	0	2	0	$0/5 = 0$	$2/8 = 0.250$	$0/4 = 0$
kasih	0	0	1	$0/5 = 0$	$0/8 = 0$	$1/4 = 0,250$
konsumsi	0	0	1	$0/5 = 0$	$0/8 = 0$	$1/4 = 0,250$
total kata	5	8	4			

Setelah menghitung frekuensi kemunculan kata dalam dokumen selanjutnya dilakukan proses perhitungan Inverse Document Frequency (IDF) untuk menentukan nilai kelangkaan kata tertentu di seluruh dokumen yang diperlihatkan pada Tabel 3.

Tabel 3. Perhitungan IDF

Kata(term)	D1	D2	D3	DF	D/DF(D=3)	IDF ($\log(D/DF)$)	IDF+1
saran	1	0	0	1	$3/1 = 3$	0.477	1.477
pkkmb	1	0	1	2	$3/2 = 1.5$	0.176	1.176
depan	1	0	1	2	$3/2 = 1.5$	0.176	1.176
lebih	1	1	0	2	$3/2 = 1.5$	0.176	1.176
seru	1	1	0	2	$3/2 = 1.5$	0.176	1.176
biar	0	1	0	1	$3/1 = 3$	0.477	1.477
tidak	0	1	0	1	$3/1 = 3$	0.477	1.477
banyak	0	1	0	1	$3/1 = 3$	0.477	1.477
bosan	0	1	0	1	$3/1 = 3$	0.477	1.477
goyang	0	1	0	1	$3/1 = 3$	0.477	1.477
kasih	0	0	1	1	$3/1 = 3$	0.477	1.477
konsumsi	0	0	1	1	$3/1 = 3$	0.477	1.477

Setelah mendapatkan nilai IDF, proses selanjutnya yakni melakukan perhitungan antara nilai TF dan IDF untuk mendapatkan hasil akhir dari proses TF-IDF untuk menggambarkan level kepentingan sebuah kata baik pada dokumen tertentu maupun di seluruh korpus yang diperlihatkan dalam Tabel 4.

Tabel 4. Perhitungan TF-IDF

Kata(term)	IDF	TF			TF*IDF		
		D1	D2	D3	D1	D2	D3
saran	1.477	0.2	0	0	0.295	0	0
pkkmb	1.176	0.2	0	0.25	0.235	0	0.294
depan	1.176	0.2	0	0.25	0.235	0	0.294
lebih	1.176	0.2	0.125	0	0.235	0.147	0
seru	1.176	0.2	0.125	0	0.235	0.147	0

Kata(term)	IDF	TF			TF*IDF		
		D1	D2	D3	D1	D2	D3
biar	1.477	0	0.125	0	0	0.185	0
tidak	1.477	0	0.125	0	0	0.185	0
banyak	1.477	0	0.125	0	0	0.185	0
bosan	1.477	0	0.125	0	0	0.185	0
goyang	1.477	0	0.25	0	0	0.369	0
kasih	1.477	0	0	0.25	0	0	0.369
konsumsi	1.477	0	0	0.25	0	0	0.369

3.6 Klasifikasi

Proses klasifikasi dikerjakan dengan memakai dua metode yakni *Naïve Bayes* dan *Random Forest*. *Naive bayes* ialah algoritma klasifikasi berbasis probabilitas yang memakai prinsip bayes untuk memperkirakan golongan dari suatu data berdasarkan fitur-fitur. Algoritma ini beranggapan bahwa tiap fitur bersifat independen dan tidak saling bergantung. *Naïve bayes* sangat efisien untuk klasifikasi teks karena kecepatan komputasi dan kemampuan dalam menangani data berdimensi tinggi, seperti vektor TF-IDF [18].

Sedangkan *random forest* yaitu algoritma ensemble yang mengkombinasikan sejumlah *decision trees* atau pohon keputusan untuk memperoleh perkiraan yang lebih valid dan konsisten [19]. Setiap pohon keputusan dilatih pada subset acak dari data atau fitur, menggunakan teknik bootstrap aggregating (bagging) dan random feature selection atau seleksi fitur random. Prediksi akhir dihasilkan dengan mengambil mayoritas suara dari semua pohon untuk tugas klasifikasi [20].

3.7 Evaluasi Model

Fase evaluasi ditujukan untuk menilai performa kedua model klasifikasi, *Naïve bayes* dan *Random Forest* dalam mengidentifikasi sentimen. Evaluasi dijalankan dengan membandingkan hasil perkiraan terhadap label aktual pada data uji. Kinerja model diukur menggunakan beberapa metrik diantaranya yakni *accuracy*, *precision*, *recall* dan *F1-score*.

Accuracy berfungsi untuk menghitung jumlah prediksi yang akurat terhadap total keseluruhan data uji. *Precision* difungsikan untuk menilai tingkat keakuratan model dalam menentukan prediksi kelas positif. *Recall* digunakan untuk mengukur kemampuan model dalam mengidentifikasi seluruh data yang tergolong dalam kelas positif. Sedangkan *F1-score* digunakan untuk mengukur rata-rata antara *precision* dan *recall* untuk memberi kestabilan antara keduanya, ketika terdapat ketidak seimbangan jumlah data tiap kelas.

4. HASIL DAN PEMBAHASAN

4.1 Preprocessing Data

Pada langkah ini, data teks ulasan mahasiswa terlebih dahulu dibersihkan untuk meningkatkan kualitas data sebelum dilakukan analisis sentimen. Proses *preprocessing* ini dilakukan melalui beberapa tahapan yaitu *case folding*, *cleansing*, *tokenization*, *normalization*, *stopword*, dan *stemming*. Hasil dari proses *preprocessing* ini ditampilkan dalam Tabel 5 yang merupakan lima data teratas.

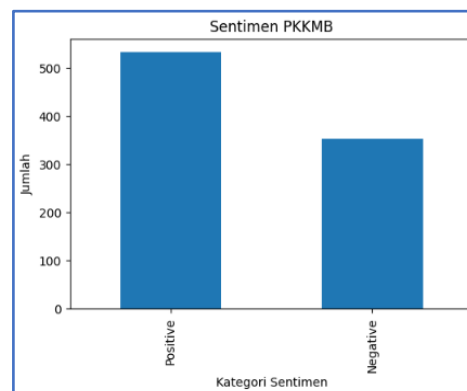
Tabel 5. Hasil Proses Preprocessing

No	Ulasan	Ulasan_lower	Ulasan_clean	Ulasan_tokenized	Ulasan_Normalizad	Ulasan_stopword	Ulasan_stemming
0	banyakin ice breaking	banyakin ice breaking	banyakin ice breaking	[banyakin, ice, breaking]	[perbanyak, ice, breaking]	[perbanyak, ice, breaking]	[banyak, ice, breaking]
	semoga panitia semakin	semoga panitia semakin	semoga panitia semakin	[semoga, panitia, semakin,	[semoga, panitia, semakin,	[semoga, panitia, kompak,	[semoga, panitia, kompak,
1	kompak dan bisa	kompak dan bisa	kompak dan bisa	kompak, dan, bisa,	kompak, dan, bisa, membuat,	pkkmb, seru]	pkkmb, seru]

	membuat pkkmb menjadi semakin seru	membuat pkkmb menjadi semakin seru	membuat pkkmb menjadi semakin seru	membuat, pkkmb, menjadi, semakin, seru]	pkkmb, menjadi, semakin, seru]		
2	jangan galak galak	jangan galak galak	jangan galak galak	[jangan, galak, galak]	[jangan, galak, galak]	[jangan, galak, galak]	[jangan, galak, galak]
3	makin seru ya kaa	makin seru ya kaa	makin seru ya kaa	[makin, seru, ya, kaa]	[makin, seru, ya, kakak]	[seru, ya, kakak]	[seru, ya, kakak]
4	adain konser kakk	adain konser kakk	adain konser kakk	[adain, konser, kakk]	[adakan, konser, kakak]	[adakan, konser, kakak]	[adakan, konser, kakak]

4.2 Labeling Data

Setelah data melewati proses preprocessing, maka langkah selanjutnya yakni proses labeling. Labeling data bertujuan untuk mengetahui persepsi mahasiswa terhadap kegiatan PKKMB baik itu persepsi yang mengandung sentimen positif maupun negatif. Dari proses labeling tersebut didapatkan sentimen positif berjumlah 533 dan sentimen negatif berjumlah 353 yang terlihat dalam grafik pada Gambar 2.



Gambar 2. Grafik sentimen

4.3 Split Data

Pembagian data dilakukan dengan fungsi *train_test_split* dari *library scikit-learn* dengan rasio 80% data latih dan 20% data uji. Proses pembagian dilakukan acak dengan parameter *random_state=42* untuk memastikan konsistensi hasil. Hasil dari proses ini, didapatkan 708 data latih dan 178 data uji.

4.4 TF-IDF

Pada proses ini teks ulasan akan diubah menjadi numerik untuk menilai tingkat kepentingan sebuah kata dalam korpus dokumen. Untuk melihat tingkat kepentingan suatu kata tersebut maka peneliti menampilkan 10 kata yang memiliki tingkat kepentingan yang terbesar secara keseluruhan pada data penelitian.

Tabel 6. 10 kata terpenting

No	Kata	TF	IDF	TF-IDF
0	lebih	0.055915	1.860073	0.104006
1	tidak	0.030529	2.793171	0.085274
2	baik	0.021674	2.909895	0.063071
3	seru	0.022456	2.696321	0.060549
4	moga	0.020639	2.711825	0.055970
5	lebih seru	0.013110	3.207147	0.042046
6	lebih baik	0.012528	3.344348	0.041899
7	semangat	0.011328	3.556522	0.040288
8	depan	0.013366	2.968736	0.039679
9	saran	0.009858	3.287189	0.032407

Berdasarkan Gambar 4, kata “lebih” menunjukkan nilai TF-IDF tertinggi (0,104006). Hal ini menunjukkan bahwa kata tersebut cukup sering muncul yang ditandai dengan nilai TF yang tinggi dan memiliki tingkat keunikan (IDF) yang relatif baik, sehingga berperan penting dalam membentuk makna ulasan.

Kata “tidak” berada di posisi kedua dengan nilai TF-IDF sebesar 0,085274. Meskipun frekuensinya tidak setinggi kata “lebih”, nilai IDF yang cukup tinggi menunjukkan bahwa kata ini bersifat informatif dan sering digunakan dalam konteks penilaian khususnya dalam menggunakan sentimen negatif.

Kata “baik”, “seru”, dan “semangat” mencerminkan ekspresi positif mahasiswa di dalam ulasan. Nilai TF-IDF yang tinggi menunjukkan bahwa sentimen positif cukup dominan dalam data, terutama dalam menggambarkan pengalaman responden.

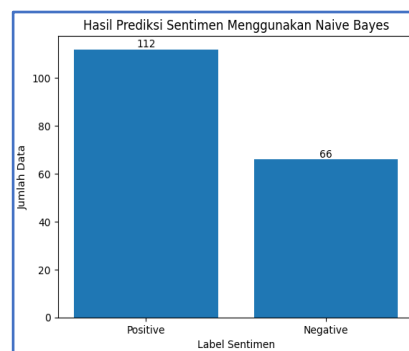
Selain itu, munculnya bigram seperti “lebih seru” dan “lebih baik” menunjukkan bahwa penggunaan n-gram mampu menangkap konteks makna yang lebih spesifik dibandingkan kata tunggal. Hal ini membuktikan bahwa kombinasi kata memberikan informasi yang lebih kuat dalam analisis sentimen.

Kata “moga”, “depan”, dan “saran” mengindikasikan adanya harapan dan masukan dari responden, yang sering digunakan dalam ulasan bersifat evaluatif.

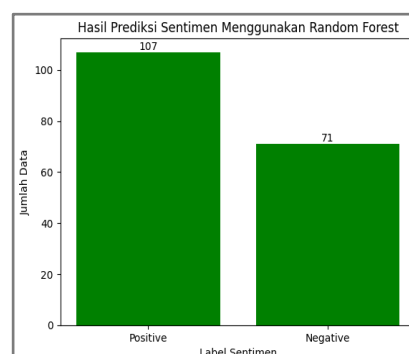
4.5 Klasifikasi

Setelah proses TF-IDF, dilakukan proses klasifikasi menggunakan Naïve Bayes dan Random Forest. Pada pengklasifikasian Naïve Bayes yang terlihat pada Gambar 3 menunjukkan Naïve Bayes dapat memprediksi sentimen positif sebanyak 112 sentimen positif dan 66 data sentimen negatif. Hasil ini memperlihatkan bahwa Naive Bayes cenderung lebih banyak mengklasifikasikan data ke dalam kelas positif. Hal tersebut disebabkan oleh karakteristik Naive Bayes yang sangat bergantung pada frekuensi kemunculan kata, sehingga lebih sensitif terhadap kata-kata dominan yang sering muncul pada kelas mayoritas.

Sedangkan Random Forest yang dapat dilihat pada Gambar 4 menunjukkan model menghasilkan prediksi sebanyak 107 data sentimen positif dan 71 data sentimen negatif. Seperti Naïve Bayes Random Forest juga memprediksi sentimen positif lebih dominan. Untuk membandingkan kedua algoritma ini perlu dilaksanakan evaluasi model terhadap keduanya.



Gambar 3. Klasifikasi Naïve Bayes



Gambar 4. Klasifikasi Random Forest

4.6 Evaluasi Model

Proses evaluasi model dilakukan untuk mengetahui kemampuan algoritma dalam mengklasifikasikan sentimen dengan benar. Pada penelitian ini, diterapkan evaluasi menggunakan *confusion matrix* sebagai dasar penghitungan metrik performa, meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* menunjukkan jumlah prediksi model yang benar maupun salah dengan membandingkan label aktual dan hasil prediksi. Matriks terdiri atas empat komponen yang dapat dilihat pada Tabel 7.

Tabel 7. *Confusion matrix*

Kelas aktual	Kelas Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan:

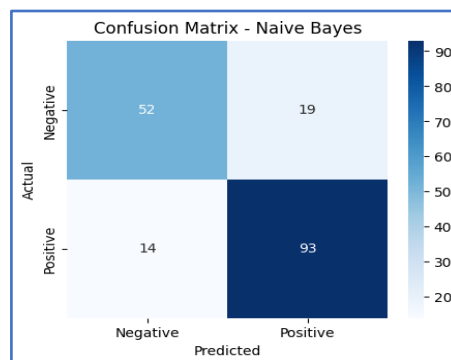
TP (True Positive) : Data dengan label yang sebenarnya positif berhasil diprediksi sebagai positif.

TN (True Negative) : Data dengan label yang sebenarnya negatif berhasil diprediksi sebagai negatif.

FP (False Positive) : Data yang sesungguhnya negatif namun salah dinilai sebagai positif.

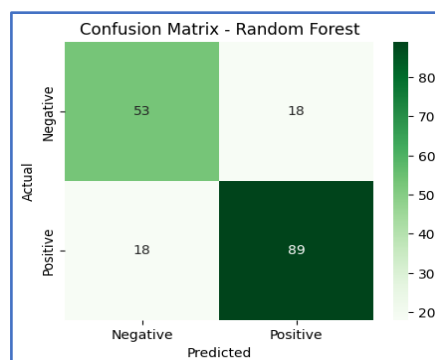
FN (False Negative) : Data yang sesungguhnya positif tetapi salah dinilai sebagai negatif.

Untuk evaluasi model Naïve bayes berdasarkan 178 data uji dapat di dilihat pada Gambar 5. Pada gambar tersebut, model Naïve Bayes berhasil mengklasifikasi 93 (TP) data positif dengan benar dan 52 data negatif dengan benar (TN), selain itu didapatkan 19 data negatif dinilai salah diprediksi sebagai positif dan 14 data positif yang salah dinilai sebagai negatif.



Gambar 5. Confusion Matrix Naïve Bayes

Lalu untuk evaluasi model random forest di dapatkan 89 data positif yang benar diperkirakan sebagai positif dan 53 data negatif benar dinilai negatif, selain itu 18 data negatif salah dinilai atau diprediksi positif dan 18 data positif salah diprediksi negatif.



Gambar 6. Confusion Matrix Random Forest

Untuk mengevaluasi kinerja model secara menyeluruh dan menentukan seberapa baik model, diperlukan sebuah pengujian. Pengujian ini dilakukan dengan menghitung *accuracy*, *precision*, *recall* dan *f1-score*. Pengukuran ini dapat diterapkan berdasarkan persamaan (4), (5), (6), dan (7).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - score = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Berdasarkan persamaan tersebut, peneliti menerapkannya pada confusion matrix yang dibuat oleh tiap model yaitu Naïve Bayes dan Random Forest. Perhitungan ini memungkinkan untuk melihat secara objektif model mana yang memiliki performa lebih baik dalam mengklasifikasikan data uji, serta aspek mana yang menjadi kekuatan atau kelemahan dari masing-masing model. Berikut penerapan *accuracy*, *Precision*, *Recall* dan *F1-Score* pada confusion matrix model Naïve Bayes dan Random Forest.

a) Pengujian pada model Naïve Bayes

$$Accuracy = \frac{93+52}{93+52+19+14} = \frac{145}{178} = 0.814 = 81.4\%$$

$$Precision = \frac{93}{93+19} = \frac{93}{112} = 0.8304 = 83.04\%$$

$$Recall = \frac{93}{93+14} = \frac{93}{107} = 0.8692 = 86.92\%$$

$$F1 - score = 2 \frac{0.8304 \times 0.8692}{0.8304 + 0.8692} = 2 \frac{0.7218}{1.6996} = 0.8493 = 84.93\%$$

b) Pengujian pada model Random Forest

$$Accuracy = \frac{89+53}{89+53+18+18} = \frac{142}{178} = 0.7978 = 79.78\%$$

$$Precision = \frac{89}{89+18} = \frac{89}{107} = 0.8318 = 83.18\%$$

$$Recall = \frac{89}{89+18} = \frac{89}{107} = 0.8318 = 83.18\%$$

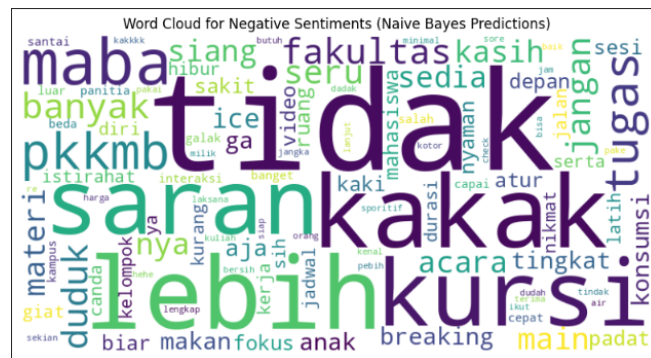
$$F1 - score = 2 \frac{0.8318 \times 0.8318}{0.8318 + 0.8318} = 0.83183 = 83.18\%$$

Berdasarkan pengujian yang telah dilakukan menunjukkan performa Naïve Bayes lebih unggul secara keseluruhan, terutama pada *accuracy*, *recall* dan *f1-score*. Model ini lebih baik untuk mengenali seluruh data positif serta menyajikan prediksi yang lebih konsisten. Sedangkan random forest sedikit lebih unggul pada *precision*, namun memiliki kelemahan pada *recall* dan akurasi. keunggulan Naive Bayes dalam penelitian ini disebabkan oleh kesesuaian algoritma dengan karakteristik data teks TF-IDF, dominasi kata-kata bermakna positif, serta kesederhanaan model yang mendukung generalisasi yang baik. Oleh karena itu, Naive Bayes terbukti lebih efektif dibandingkan Random Forest dalam melakukan klasifikasi sentimen pada dataset evaluasi PKKMB

di Universitas Merdeka Malang, dengan nilai accuracy 81.4%, precision 83.04%, recall 86.92% dan f1-score 84.93%

Tabel 8. Kata dan frekuensi sentimen negatif

Kata	Frekuensi	Kata	Frekuensi
Tidak	121	Seru	20
Lebih	69	Panitia	18
saran	50	Acara	16
kakak	45	Tugas	16
Jangan	38	Fakultas	16
Pkkmb	32	Tolong	15
Duduk	28	Tingkat	15
Materi	24	depan	14
Maba	21	Konsumsi	13
kursi	21	Kurang	13



Gambar 7. *Wordcloud* sentimen negatif

Berdasarkan Tabel 8 dan Gambar 7, terlihat bahwa kata “tidak” memiliki frekuensi tertinggi, yang menunjukkan dominasi ketidakpuasan mahasiswa baru terhadap pelaksanaan PKKMB. Tingginya frekuensi kata “lebih” dan “saran” mengindikasikan adanya ekspektasi peserta terhadap peningkatan kualitas kegiatan. Aspek interaksi penyelenggara tercermin dari kemunculan kata “kakak” dan “panitia”, yang menunjukkan bahwa komunikasi dan pola pendampingan masih perlu dievaluasi.

Dari sisi kenyamanan, kata “duduk” dan “kursi” menegaskan bahwa durasi kegiatan serta fasilitas pendukung menjadi sumber keluhan utama. Selain itu, kata “materi” dan “tugas” menunjukkan perlunya penyesuaian beban dan metode penyampaian kegiatan. Oleh karena itu, hasil analisis ini digunakan sebagai dasar rekomendasi perbaikan PKKMB, meliputi pengelolaan waktu yang lebih efektif, peningkatan kenyamanan fasilitas, penguatan pendekatan humanis panitia, serta perancangan kegiatan yang lebih interaktif dan proporsional sesuai kebutuhan mahasiswa baru.

5. KESIMPULAN

Berdasarkan hasil penelitian dan analisis yang dilakukan, dapat disimpulkan bahwa algoritma Naïve Bayes lebih unggul dibandingkan Random Forest pada dataset evaluasi kegiatan PKKMB berdasarkan pelabelan otomatis menggunakan IndoBERT. Keunggulan tersebut tercermin dari nilai accuracy, recall dan f1-score yang tinggi, yang mengindikasikan bahwa Naïve Bayes mampu mengkategorikan sentimen mahasiswa secara tetap dan efisien. Hal ini menunjukkan bahwa Naïve Bayes merupakan algoritma yang tepat untuk diterapkan pada data teks evaluasi pendidikan dengan karakteristik serupa.

Dari sisi evaluasi program, hasil analisis sentimen menunjukkan bahwa mayoritas tanggapan mahasiswa terhadap pelaksanaan PKKMB bersifat positif, yang mencerminkan tingkat penerimaan dan kemasakan yang baik terhadap kegiatan tersebut. Meskipun demikian, ditemukan pula sejumlah umpan balik negatif yang perlu mendapat perhatian, khususnya terkait durasi kegiatan yang dianggap terlalu panjang serta metode penyampaian materi yang kurang interaktif. Temuan ini

memberikan dasar yang jelas dan berbasis data bagi pihak universitas untuk melakukan evaluasi dan perbaikan program secara terarah.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan analisis sentimen berbasis machine learning dengan pelabelan otomatis dapat dimanfaatkan sebagai alat pendukung dalam pengambilan keputusan di bidang pendidikan. Dengan memanfaatkan hasil klasifikasi sentimen, institusi pendidikan dapat memperoleh gambaran objektif mengenai persepsi mahasiswa serta menyusun strategi peningkatan kualitas kegiatan secara berkelanjutan.

DAFTAR PUSTAKA

- [1] Universitas Merdeka Malang, *Panduan Umum Pengenalan Kehidupan Kampus Mahasiswa Baru Universitas Merdeka Malang*. 2025.
- [2] I. Azizah, I. Cholissodin, and N. Yudistira, "Analisis Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play menggunakan Metode Word Embedding dan Long Short Term Memory (LSTM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2453–2459, 2023.
- [3] R. C. Rivaldi and T. D. Wismarini, "Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP)," *Elkom J. Elektron. dan Komput.*, vol. 17, no. 1, pp. 120–128, 2024, doi: 10.51903/elkom.v17i1.1680.
- [4] T. F. Basar, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest," vol. 6, no. 3, pp. 1426–1433, 2022.
- [5] A. Miftahusalam, H. Pratiwi, and I. Slamet, "Perbandingan Metode Random Forest dan Naive Bayes pada Analisis Sentimen Review Aplikasi BCA Mobile," *SIPTEK Semin. Nas. Inov. dan Pengemb. Teknol. Pendidik.*, vol. 1, no. 1, pp. 1–8, 2023, doi: 10.59395/ijadis.v5i2.1342.
- [6] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 563–572, 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.
- [7] M. Yasir and R. Suraji, "Perbandingan Metode Klasifikasi Naïve Bayes , Decision Tree , Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube," *J. Cahaya Mandalika*, vol. 3, no. 2, pp. 180–192, 2023.
- [8] Z. Azfil Yugian, A. Hananto, T. Tukino, and E. Novalias, "Klasifikasi Data Teks Ulasan Toko Sepatu Sovella Di Shopee Menggunakan Algoritma Naive Bayes Dalam Proses Data Mining," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 5, pp. 8732–8736, 2025, doi: 10.36040/jati.v9i5.15159.
- [9] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," vol. 6, no. 9, pp. 4305–4313, 2022.
- [10] M. P. . Putra and K. R. . Wardani, "Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial," *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 05, no. 01, pp. 63–71, 2020.
- [11] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," *J. Inf. Syst. public Heal.*, vol. 4, no. 1, pp. 28–36, 2019.
- [12] A. A. Syam, G. Hardy M, A. Salim, D. F. Surianto, and M. Fajar B, "Analisis Teknik Preprocessing Pada Sentimen Masyarakat Terkait Konflik Israel-Palestina Menggunakan Support Vector Machine," *JIFI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1464–1472, 2024, doi: 10.29100/jifi.v9i3.5527.
- [13] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknol.)*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.
- [14] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine

- Learning,” vol. 9, no. April, pp. 19–28, 2023, doi: 10.34128/jsi.v9i1.622.
- [15] R. Darmawan and S. Amini, “Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor pada Twitter Comparison of Sentiment Analysis Results Using Naïve Bayes and K-Nearest Neighbor Algorithm on Twitter,” *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 495–501, 2022, [Online]. Available: <https://senafiti.budiluhur.ac.id/index.php/>
- [16] W. S. Hoar, A. Zubair, and L. Muflikhah, “Analisis sentimen kebijakan masuk sekolah pukul lima pagi menggunakan algoritma Naïve Bayes,” *J. Inf. Syst. Appl. Dev.*, vol. 2, no. 1, pp. 20–30, 2024, doi: 10.26905/jisad.v2i1.10935.
- [17] A. Sabrani, I. G. W. Wedashwara W., and F. Bimantoro, “Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia,” *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 2, no. 1, pp. 89–100, 2020, doi: 10.29303/jtika.v2i1.87.
- [18] J. D. P. Beneng and A. Zubair, “Analisis Sentimen Pengguna Twitter terhadap AI-Generated Art Menggunakan Metode Naive Bayes dan Neural Network Classifier,” *Semin. Nas. Sist. Inf.*, vol. 8, pp. 51–61, 2024, [Online]. Available: <https://jurnalfti.unmer.ac.id/index.php/senasif/article/view/636>
- [19] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes , Random Forest Dan Support Vector Machine,” vol. 18, no. 1, pp. 71–80, 2020.
- [20] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021, doi: 10.26877/jiu.v7i1.7099.