COMPUTA
Journal of Computer and Informatics Engineering *p-ISSN: XXX-XXXX*
Volume 2, Nomor 1, 2026 *e-ISSN: XXXX-XXX*

ANALISIS SENTIMEN KAMPUS MENGAJAR MENGGUNAKAN PERBANDINGAN METODE *NAIVE BAYES* DENGAN *SUPPORT VECTOR MACHINE*

Ahmad Syifa[1)], Indra Maulana[2)], Diah Afrianti Rahayu[3)]

[1)2)3)]Institut Prima Bangsa Cirebon
[1)]syifaahmad896@gmail.com, [2)]indramaulana360@gmail.com, [3)]drahyu@ipbcirebon.ac.id

Abstrak

Program Kampus Mengajar, sebagai bagian dari kebijakan Merdeka Belajar Kampus Merdeka (MBKM), telah melibatkan ribuan mahasiswa dalam kegiatan pengajaran di berbagai daerah di Indonesia, sehingga memunculkan beragam opini yang perlu dipahami secara sistematis oleh penyelenggara program. Penelitian ini bertujuan untuk mengidentifikasi kecenderungan sentimen mahasiswa terhadap program tersebut sekaligus membandingkan kinerja dua algoritma klasifikasi teks, yaitu *Naive Bayes* dan *Support Vector Machine (SVM),* dalam mengklasifikasikan opini menjadi sentimen positif dan negatif. Data penelitian dikumpulkan melalui kuesioner terhadap 81 mahasiswa peserta Kampus Mengajar (IPB angkatan 5–8, UGJ angkatan 6, dan mahasiswa penempatan wilayah Cirebon), kemudian diberi label secara manual dan dibagi dengan rasio 70:30 untuk data latih dan data uji. Data diproses melalui tahapan *text preprocessing (cleansing, tokenizing, case folding, filtering, dan stopword removal)* sebelum diklasifikasikan menggunakan RapidMiner. Hasil penelitian menunjukkan bahwa dari 81 data, 61 di antaranya (75,3%) menunjukkan sentimen positif dan 20 data (24,7%) menunjukkan sentimen negatif, dengan opini positif didominasi peningkatan soft skill, kepercayaan diri, dan pengalaman langsung di dunia pendidikan, sementara opini negatif berkaitan dengan beban administrasi, minimnya bimbingan lapangan, dan kendala koordinasi program. Pada tahap klasifikasi, algoritma SVM mencapai akurasi tertinggi sebesar 86,11% dengan presisi sentimen positif 100% namun *recall* 78,26%, sedangkan *Naive Bayes* mencapai akurasi 83,33% dengan performa presisi dan *recall* yang lebih seimbang (presisi positif 77,7%, *recall* positif 87,5%). Temuan ini membuktikan bahwa pemilihan algoritma klasifikasi berpengaruh signifikan terhadap akurasi maupun kestabilan hasil analisis sentimen, sehingga hipotesis alternatif (H_1) dalam penelitian ini diterima. Penelitian ini memberikan kontribusi ganda: secara akademik, hasil ini menjadi acuan metodologis bagi pengembangan analisis sentimen berbasis machine learning di sektor pendidikan, khususnya dalam memilih algoritma yang tepat sesuai karakteristik dataset opini; secara praktis, temuan ini dapat dimanfaatkan oleh penyelenggara Program Kampus Mengajar sebagai dasar evaluasi untuk memperbaiki aspek supervisi lapangan, beban administratif, dan pengalaman mahasiswa, sekaligus menjadi bukti bahwa pendekatan berbasis survei digital dan machine learning mampu menghasilkan evaluasi program pendidikan yang lebih objektif, efisien, dan terukur.

Kata kunci: Analisis Sentimen, Kampus Mengajar, *Naive Bayes*, *Support Vector Machine*, Klasifikasi Teks

Abstract

The Kampus Mengajar Program, as part of the Merdeka Belajar Kampus Merdeka (MBKM) policy, has involved thousands of university students in teaching activities across various regions of Indonesia, resulting in diverse opinions that need to be systematically understood by program

DOI: xxxxxxx
Received: 31-08-2026
Accepted: 08-05-2026

organizers. This study aims to identify the sentiment tendencies of students toward the program while comparing the performance of two text classification algorithms, namely Naive Bayes and Support Vector Machine (SVM), in classifying opinions into positive and negative sentiments. The research data were collected through questionnaires administered to 81 Kampus Mengajar participants, consisting of students from IPB cohorts 5–8, UGJ cohort 6, and students assigned to the Cirebon region. The data were manually labeled and divided using a 70:30 ratio for training and testing data. The data were processed through text preprocessing stages, including cleansing, tokenizing, case folding, filtering, and stopword removal, before being classified using RapidMiner. The results showed that out of 81 data points, 61 (75.3%) expressed positive sentiments, while 20 (24.7%) expressed negative sentiments. Positive opinions were predominantly associated with improvements in soft skills, self-confidence, and direct experience in the field of education, while negative opinions were related to administrative burdens, limited field supervision, and program coordination issues. In the classification stage, the SVM algorithm achieved the highest accuracy of 86.11%, with a positive sentiment precision of 100% and recall of 78.26%, whereas Naive Bayes achieved an accuracy of 83.33% with a more balanced performance, obtaining a positive sentiment precision of 77.7% and recall of 87.5%. These findings demonstrate that the selection of a text classification algorithm significantly affects the accuracy and stability of sentiment analysis results; therefore, the alternative hypothesis (H_1) is accepted. This study provides a dual contribution: academically, the findings serve as a methodological reference for the development of machine learning-based sentiment analysis in the education sector, particularly in selecting appropriate algorithms according to the characteristics of opinion datasets; practically, the findings can be utilized by Kampus Mengajar program organizers as a basis for evaluating and improving field supervision, administrative workload, and students' overall program experience. Furthermore, the study demonstrates that a digital survey-based approach combined with machine learning can provide a more objective, efficient, and measurable evaluation of educational programs.

Keywords: *Sentiment Analysis, Kampus Mengajar, Naive Bayes, Support Vector Machine, Text Classification.*

1. PENDAHULUAN

Pendidikan merupakan salah satu aspek paling fundamental dalam kehidupan manusia karena memiliki kemampuan mengubah individu, termasuk mendorong mobilitas strata sosial melalui akses yang merata bagi setiap orang. Untuk mencapai tujuan pendidikan nasional, diperlukan sistem yang terpadu guna meningkatkan vitalitas intelektual bangsa sekaligus mempromosikan keadilan sosial di bidang Pendidikan [12]. Dalam kerangka inilah berbagai kebijakan dan program inovatif terus dikembangkan oleh pemerintah untuk menjembatani kesenjangan pendidikan di berbagai wilayah Indonesia.

Salah satu wujud nyata dari upaya tersebut adalah Program Kampus Mengajar, yang diinisiasi melalui kebijakan Merdeka Belajar Kampus Merdeka (MBKM). Program ini memberikan kesempatan bagi mahasiswa untuk terlibat langsung dalam kegiatan pengajaran di sekolah dasar, baik di wilayah perkotaan maupun pedesaan. Selain berkontribusi terhadap pemerataan kualitas pendidikan, program ini juga dirancang untuk mengasah soft skill dan hard skill mahasiswa, sekaligus mempersiapkan mereka menjadi calon pemimpin masa depan yang adaptif terhadap perubahan zaman [2].

Sebagai program yang melibatkan ribuan mahasiswa dengan kondisi penempatan yang sangat beragam, Kampus Mengajar secara alami memunculkan opini yang bervariasi mulai dari pengalaman positif berupa pengembangan diri, hingga keluhan terkait teknis pelaksanaan di

lapangan. Keberagaman opini inilah yang menjadi masalah inti yang perlu dijawab bagaimana memahami secara sistematis dan objektif kecenderungan sentimen mahasiswa terhadap program ini, mengingat opini yang tersebar luas melalui survei maupun media sosial sulit dipetakan secara manual dalam jumlah besar. Analisis sentimen menjadi pendekatan yang relevan karena mampu mengklasifikasikan opini ke dalam kategori positif dan negatif secara lebih cepat dan konsisten dibandingkan pembacaan manual [15].

Analisis sentimen sendiri merupakan bagian dari *text mining* yang memanfaatkan teknik komputasi untuk mengenali opini serta respons emosional individu terhadap suatu subjek [12]. Sentimen dipahami sebagai ekspresi individu yang mencerminkan perspektif seseorang terhadap suatu hal [5], sementara analisis sentimen secara khusus berfokus pada penelusuran pendapat pribadi yang mengandung kecenderungan positif maupun negatif terhadap suatu isu, produk, atau kebijakan [11]. Pendekatan berbasis machine learning menjadi krusial dalam konteks ini karena memungkinkan pengolahan dataset opini berskala besar secara akurat dan efisien.

Dalam praktiknya, terdapat berbagai algoritma klasifikasi teks yang umum digunakan dalam analisis sentimen, dua di antaranya yang paling banyak dibandingkan adalah *Naive Bayes* dan *Support Vector Machine (SVM)*. *Naive Bayes* dikenal efisien dan mudah diimplementasikan untuk pemrosesan teks, meskipun memiliki kelemahan pada asumsi independensi fitur yang tidak selalu terpenuhi pada data riil [3]. Di sisi lain, SVM bekerja dengan mencari hyperplane pemisah optimal antar kelas dan unggul dalam menangani data berdimensi tinggi [4]. Meskipun kedua metode ini kerap dibandingkan pada berbagai domain seperti perdagangan elektronik dan media sosial, penelitian yang secara khusus mengaplikasikan dan membandingkan keduanya pada konteks evaluasi program pendidikan seperti Kampus Mengajar masih sangat terbatas — inilah celah penelitian *(research gap)* yang mendasari studi ini.

Sebagai gambaran awal atas fenomena yang terjadi, observasi terhadap 81 tanggapan mahasiswa peserta Kampus Mengajar yang dikumpulkan pada 15 Mei 2025 menunjukkan distribusi sentimen sebagai berikut:

Tabel 1. Data Observasi Awal

No	Kategori Sentimen	Jumlah Data	Persentase
1	Positif	61	75,3%
2	Negatif	20	24,7%
Total		**81**	**100%**

Data awal ini memperlihatkan bahwa meskipun mayoritas mahasiswa memberikan tanggapan positif terutama terkait pengembangan soft skill dan pengalaman langsung di dunia Pendidikan proporsi tanggapan negatif yang mencapai hampir seperempat dari keseluruhan data tetap signifikan untuk ditelusuri lebih jauh, khususnya keluhan seputar beban administrasi, minimnya bimbingan lapangan, dan kendala koordinasi program.

Urgensi penelitian ini muncul dari kebutuhan mendesak untuk memahami secara akurat persepsi mahasiswa terhadap Kampus Mengajar, sebuah program strategis Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi yang tergolong masih baru dan berdampak langsung terhadap proses pembelajaran di sekolah-sekolah penempatan. Tanpa evaluasi yang sistematis, potensi masalah di lapangan seperti ketimpangan bimbingan atau beban administratif berlebih berisiko terus berulang pada periode-periode program berikutnya. Oleh karena itu, perbandingan kinerja algoritma *Naive Bayes* dan *Support Vector Machine* dalam mengklasifikasikan sentimen mahasiswa bukan hanya berkontribusi pada pengembangan metodologi analisis data teks di sektor pendidikan secara ilmiah, tetapi juga memberikan manfaat

praktis langsung bagi penyelenggara program dalam merumuskan langkah perbaikan yang lebih tepat sasaran dan berbasis data.

2. TINJAUAN PUSTAKA

2.1 Kampus Mengajar

Kampus Mengajar merupakan salah satu program unggulan dalam kebijakan Merdeka Belajar Kampus Merdeka (MBKM) yang bertujuan meningkatkan keterampilan mahasiswa sekaligus mendorong pemerataan kualitas pendidikan dasar melalui partisipasi dan kontribusi langsung mahasiswa selama satu semester penugasan [10]. Program ini menugaskan mahasiswa untuk mengajar di sekolah dasar, baik di wilayah perkotaan maupun pedesaan, sekaligus menghadapkan mereka pada berbagai tantangan lapangan yang menguji kreativitas, inovasi, kepemimpinan, komunikasi, kemampuan pemecahan masalah, dan kerja sama tim [2]. Sebagai program berskala nasional yang melibatkan ribuan peserta dari berbagai latar belakang dan wilayah penempatan, Kampus Mengajar secara alami menghasilkan pengalaman yang beragam, sehingga opini serta persepsi mahasiswa terhadap pelaksanaannya menjadi topik yang relevan untuk dikaji secara sistematis, khususnya melalui pendekatan analisis data berbasis teks.

2.2 Analisis Sentimen

Analisis sentimen, atau pengolahan opini (opinion mining), adalah pendekatan yang digunakan untuk mengidentifikasi, menggali, dan menginterpretasikan data berbasis teks secara otomatis guna mengetahui sikap, emosi, atau kecenderungan yang tersirat dalam suatu pernyataan opini, yang prosesnya dapat dilakukan dengan bantuan metode pembelajaran mesin [17]. Analisis Sentimen dapat didefinisikan sebagai kajian tentang bagaimana individu berpikir dan bereaksi melalui pendapat, pandangan, dan emosi mereka. Sebagai bagian dari data mining, analisis sentimen berfungsi menggali informasi berharga dari teks terkait produk, layanan, organisasi, tokoh, maupun program tertentu [7]. Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengklasifikasikan opini mahasiswa terhadap Program Kampus Mengajar ke dalam dua kategori, yaitu sentimen positif dan negatif, sebagai dasar evaluasi program.

2.3 Text Mining

Text mining, atau penambangan teks, merupakan cabang dari penambangan data (data mining) yang berfokus pada analisis teks dari berbagai jenis dokumen untuk mengidentifikasi informasi yang relevan serta memahami keterkaitan antar dokumen yang dikaji [1]. Salah satu area penerapan utama text mining adalah analisis sentimen, yang bertujuan mengungkap bagaimana seseorang mengekspresikan pendapat, perasaan, dan emosinya dalam bentuk teks, kemudian mengklasifikasikannya menjadi pernyataan positif atau negatif [9]. Dalam alur penelitian ini, text mining menjadi kerangka metodologis yang menaungi keseluruhan proses, mulai dari pengumpulan data opini mentah hingga transformasinya menjadi data terstruktur yang siap diklasifikasikan oleh algoritma machine learning.

2.4 Preprocessing

Preprocessing merupakan tahapan awal dalam pengolahan data teks yang bertujuan menghilangkan elemen-elemen yang tidak diperlukan serta mengubah teks mentah yang belum terstruktur menjadi bentuk yang lebih rapi dan siap diolah [13]. Tahapan ini krusial karena kualitas data yang telah dibersihkan akan sangat memengaruhi akurasi model klasifikasi yang dihasilkan. Terdapat beberapa sub-tahapan yang umum digunakan, yaitu:

Tabel 2. Tahapan *Preprocessing*

No	Tahapan	Fungsi
1	*Cleansing*	Menghapus atribut yang tidak bermakna seperti simbol, tanda baca, emoji, HTML, dan tagar (#).
2	*Case folding*	Menyeragamkan seluruh huruf menjadi huruf kecil (*lowercase*) agar sistem tidak salah membedakan kata yang identik.
3	*Tokenize*	Memecah kalimat menjadi unit-unit kecil berupa kata (*token*) untuk mempermudah analisis.
4	*Filtering*	Menyaring token berdasarkan jumlah huruf, misalnya membuang kata yang terlalu pendek atau terlalu panjang.
5	*Stopword removal*	Menghapus kata penghubung atau kata depan yang tidak memiliki makna signifikan, seperti "yang", "dan", "di", "ke".

Kelima tahapan ini dijalankan secara berurutan sebelum data opini diklasifikasikan menggunakan algoritma machine learning, dengan tujuan mengurangi noise dan meningkatkan efektivitas proses klasifikasi.

2.5 Confusion Matrix

Confusion matrix (matriks kebingungan) adalah alat evaluasi yang digunakan untuk menilai baik-buruknya kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap nilai aktual. Tabel ini mengelompokkan hasil ke dalam empat kategori: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN), sebagaimana disajikan berikut:

Tabel 3. Kalsifikasi *Confusion Matrix*

Prediksi / Aktual	Positif	Negatif
Positif	*TP*	FP
Negatif	*FN*	TN

Dari keempat komponen tersebut, tiga metrik evaluasi utama dapat dihitung :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Presicion = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Accuracy menggambarkan proporsi prediksi yang benar secara keseluruhan, *precision* menunjukkan ketepatan model dalam memprediksi kelas positif, sedangkan *recall* mengukur kemampuan model dalam menemukan seluruh data positif yang sebenarnya. Ketiga metrik ini menjadi dasar untuk membandingkan performa algoritma *Naive Bayes* dan *Support Vector* Machine dalam penelitian ini.

2.6 Naïve Bayes

Algoritma *Naive Bayes* merupakan metode klasifikasi yang didasarkan pada Teorema Bayes, dengan asumsi "naif" bahwa setiap fitur atau atribut dalam data bersifat independen satu sama lain [3]. Metode ini banyak digunakan dalam pembelajaran mesin karena kesederhanaannya, implementasinya yang cepat, serta kemampuannya mengolah data berskala besar secara efisien [6]. Dalam proses klasifikasi, *Naive Bayes* memanfaatkan probabilitas kemunculan kata pada data historis untuk memprediksi kelas suatu data baru [9]. Meskipun demikian, algoritma ini memiliki

keterbatasan pada asumsi independensi fitur yang tidak selalu terpenuhi dalam data teks di dunia nyata, yang berpotensi memengaruhi akurasi hasil klasifikasi.

2.7 Support Vector Machine

Support Vector Machine (SVM) adalah algoritma klasifikasi yang bekerja dengan mencari garis pemisah optimal, yang disebut hyperplane, untuk memisahkan data ke dalam dua kelas berbeda, yaitu positif dan negatif [8]. Pemisahan ini dapat dilakukan secara linier maupun melalui pendekatan non-linier dengan memanfaatkan fungsi kernel pada ruang berdimensi tinggi, sehingga SVM mampu menangani data yang memiliki jumlah variabel lebih besar dibandingkan jumlah observasinya [13]. Karakteristik inilah yang membuat SVM sering kali menghasilkan akurasi klasifikasi yang tinggi, khususnya pada data teks dengan dimensi fitur yang besar, meskipun umumnya membutuhkan proses komputasi yang lebih kompleks dibandingkan *Naive Bayes.*

3. METODE PENELITIAN

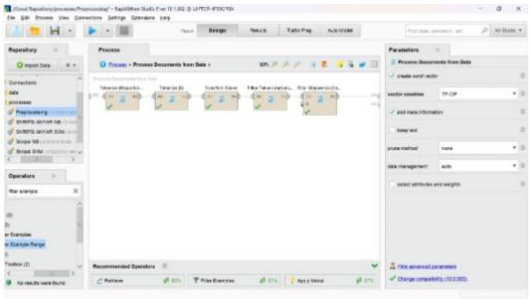
Penelitian ini menggunakan metode kuantitatif dengan pendekatan komparatif, yang bertujuan untuk membandingkan kinerja dua algoritma klasifikasi teks, yaitu *Naive Bayes* (NB) dan Support Vector Machine (SVM), dalam menganalisis sentimen mahasiswa terhadap Program Kampus Mengajar. Variabel independen dalam penelitian ini adalah algoritma NB (x_1) dan SVM (x_2), sedangkan variabel dependennya adalah hasil klasifikasi berupa sentimen positif (y_1) dan sentimen negatif (y_2). Penelitian dilaksanakan pada periode April–Mei 2025 dengan menyebarkan kuesioner kepada mahasiswa di lingkungan Kota dan Kabupaten Cirebon.

Populasi penelitian terdiri atas 171 mahasiswa peserta Program Kampus Mengajar, mencakup mahasiswa Institut Prima Bangsa (IPB) Cirebon angkatan 5–8 sebanyak 65 orang, mahasiswa yang ditempatkan di wilayah Cirebon sebanyak 106 orang, serta mahasiswa Universitas Swadaya Gunung Jati (UGJ) Cirebon angkatan 5–7. Teknik pengambilan sampel yang digunakan adalah purposive sampling, yaitu penentuan sampel berdasarkan kriteria tertentu dalam hal ini mahasiswa yang pernah mengikuti dan memiliki pengalaman langsung terhadap program tersebuT dengan target responden berkisar 60 hingga 100 mahasiswa. Data dikumpulkan melalui dua teknik, yaitu observasi menggunakan kuesioner yang disebarkan melalui grup WhatsApp dan pengiriman langsung kepada mahasiswa, serta studi literatur untuk memperkaya referensi mengenai algoritma NB dan SVM.

Analisis data dalam penelitian ini dilakukan melalui tiga tahapan utama, yaitu preprocessing, penerapan algoritma, dan evaluasi model, dengan penjelasan sebagai berikut:

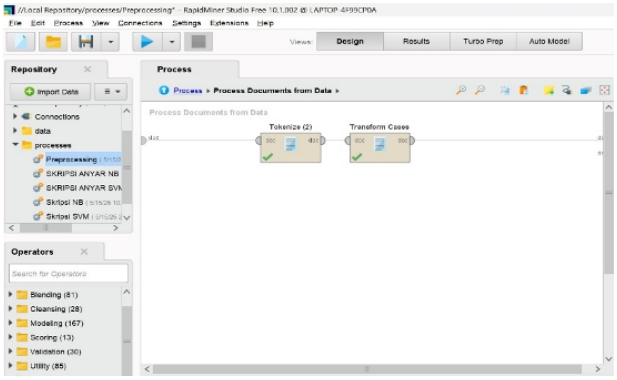
3.1 Tahap Preprocessing Data

Opini mentah dari kuesioner terlebih dahulu melalui proses preprocessing menggunakan RapidMiner untuk membersihkan dan menyeragamkan teks sebelum diklasifikasikan, melalui lima sub-tahapan berikut:



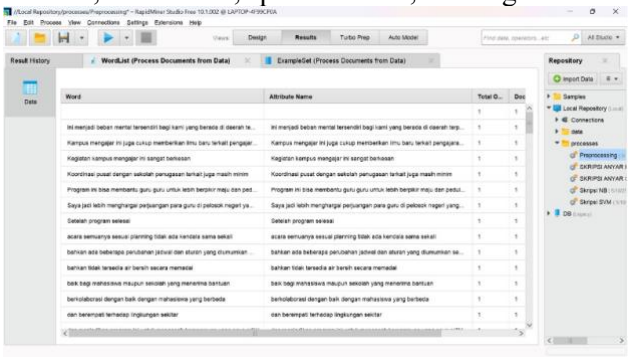
Gambar 1. Tahap Preprocessing

3.1.1 ***Case folding***, menyeragamkan seluruh huruf menjadi huruf kecil (*lowercase*) agar variasi penulisan huruf besar/kecil tidak dianggap berbeda oleh sistem.



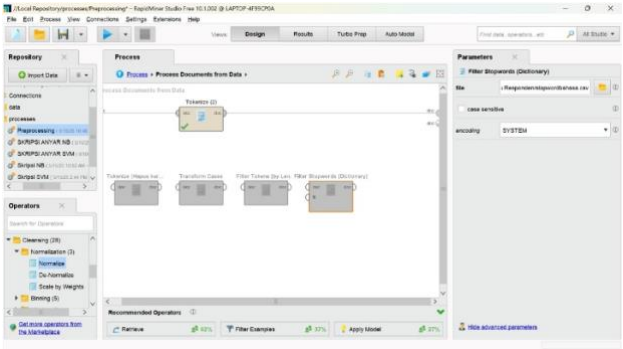
Gambar 2. Prosess Case Folding

3.1.2 ***Cleansing***, menghapus karakter atau elemen yang tidak bermakna seperti email, simbol "@", URL, emotikon, tanda baca, spasi berlebih, dan angka.



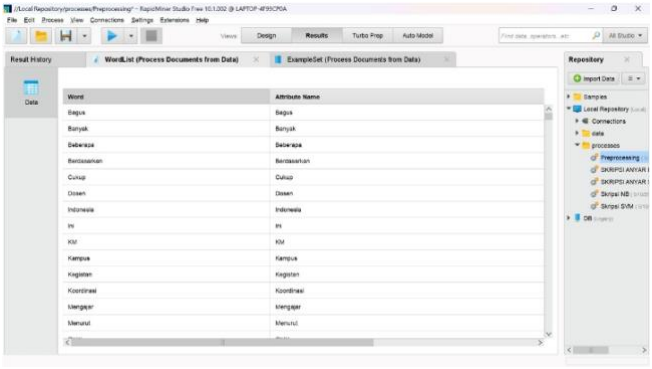
Gambar 3. Hasil Cleansing

3.1.3 ***Tokenize***, memecah kalimat menjadi satuan kata (token) agar teks dapat dianalisis per kata; pada tahap ini dihasilkan 442 token dari keseluruhan data.



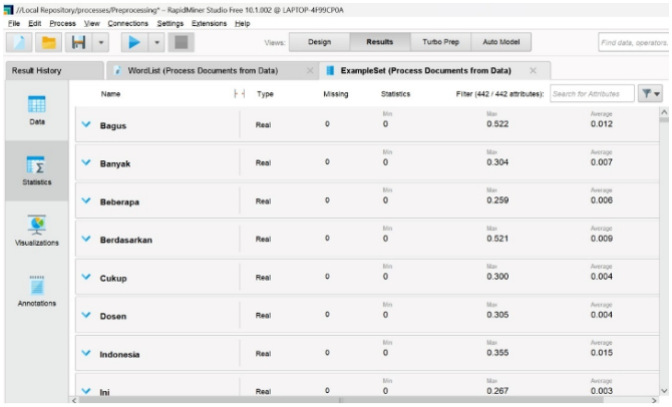
Gambar 4. Proses Tokenize

Hasil dari proses Tokenize, dapat dilihat pada gambar 5. Hasil tokenize yang menunjukkan token dari setiap kalimat.



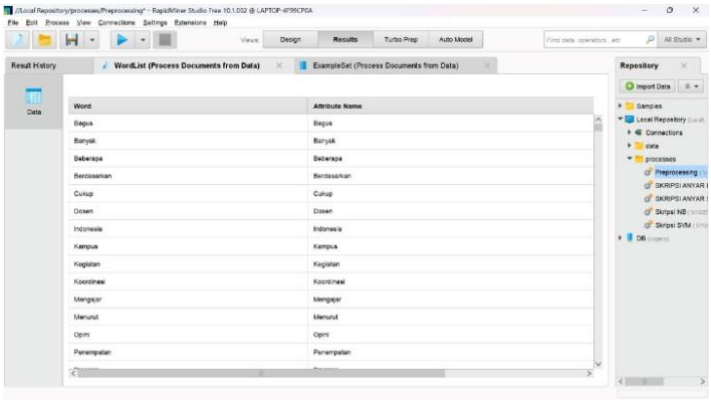
Gambar 5. Hasil Tokenize

Pada Gambar 6. Dapat dilihat setelah dilakukan proses tokenize, jumlah kata yang dihasilkan pada proses ini sebanyak 442 kata.



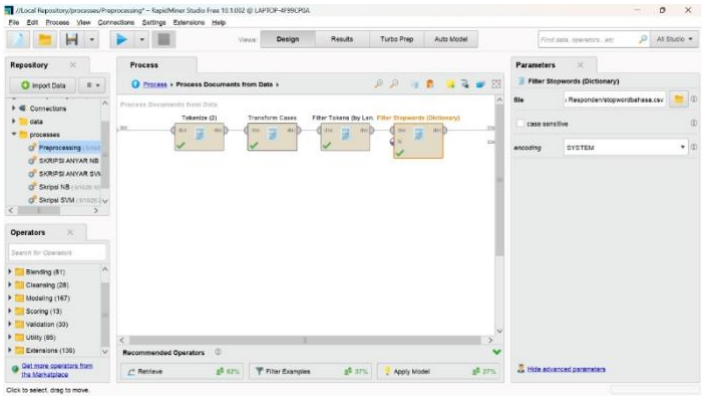
Gambar 6. Jumlah token setelah dilakukan tokenize

3.1.4 ***Filtering***, menyaring token dengan membuang kata yang memiliki kurang dari 4 huruf atau lebih dari 25 huruf, untuk mengurangi kata yang tidak relevan.



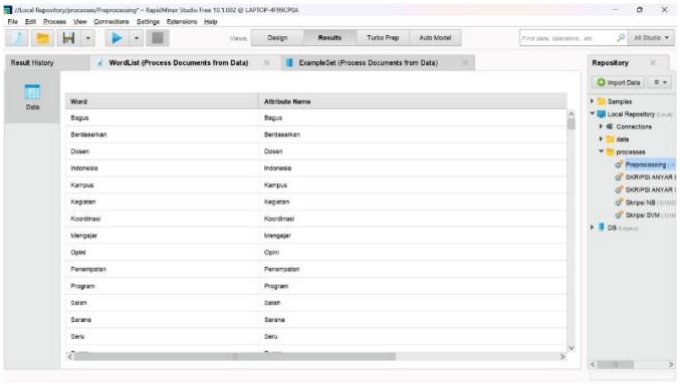
Gambar 7. Hasil Filtering

3.1.5 ***Stopword removal***, menghapus kata-kata yang tidak memiliki nilai informasi penting seperti kata sambung, kata depan, kata ganti, dan kata bantu, menggunakan daftar stoplist berisi 758 kata yang diperoleh dari Kaggle.



Gambar 8. Proses Stopword

Setelah memasukan stoplist kedalam operator filter stopword maka selanjutnya data akan di proses dan menghasilkan kata yang sudah disaring sehingga hanya tersisa kata tidak tercantum didalam stoplist tersebut. Gambar 9 menunjukkan hasil dari proses Stopword.



Gambar 9. Hasil Stopword

Setelah melalui seluruh tahapan tersebut, jumlah kata dalam data berkurang secara bertahap, dari 442 kata pada tahap tokenize menjadi 425 kata setelah case folding, dan terus berkurang setelah filtering serta stopword removal, sehingga data siap untuk tahap klasifikasi.

3.2 Tahap Penerapan Algoritma

Tahap Penerapan Algoritma Data yang telah bersih kemudian dibagi dengan rasio 70% data latih dan 30% data uji, lalu diklasifikasikan menggunakan dua algoritma:

3.2.1 ***Naive Bayes***, yang bekerja berdasarkan Teorema Bayes dengan menghitung peluang suatu data termasuk ke dalam kelas tertentu berdasarkan probabilitas kemunculan kata pada data historis, menggunakan persamaan:

$$\boldsymbol{P(Ci|X)} = \frac{\boldsymbol{P(X|Ci)} \cdot \boldsymbol{P\,(Ci)}}{\boldsymbol{P\,(X)}}$$

dengan asumsi bahwa setiap fitur (kata) bersifat independen satu sama lain.

3.2.2 ***Support Vector Machine***, yang bekerja dengan mencari *hyperplane* pemisah optimal antara kelas positif dan negatif melalui persamaan:

$$f\,(x) = w \cdot x + b$$

di mana w dan b merupakan parameter hyperplane yang dicari agar jarak (margin) antara kedua kelas data menjadi maksimal.

3.3 Tahap Evaluasi Model

Kinerja kedua algoritma dievaluasi menggunakan confusion matrix, yang membandingkan hasil prediksi model dengan label aktual ke dalam empat kategori: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Dari confusion matrix tersebut dihitung tiga metrik utama:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Presicion = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Accuracy menunjukkan proporsi prediksi yang benar secara keseluruhan, precision mengukur ketepatan model dalam memprediksi kelas positif, dan recall mengukur kemampuan model menemukan seluruh data positif yang sebenarnya ada. Ketiga metrik ini digunakan untuk membandingkan mana di antara *Naive Bayes* dan SVM yang memberikan performa klasifikasi lebih baik dan lebih stabil.

Secara keseluruhan, prosedur penelitian ini dilaksanakan melalui enam tahapan berurutan:

3.3.1 **Persiapan,** meliputi perancangan desain penelitian dan instrumen kuesioner;
3.3.2 **Pengumpulan data**, yaitu penyebaran kuesioner kepada mahasiswa peserta Kampus Mengajar;
3.3.3 **Pengolahan data**, melalui tahapan preprocessing seperti dijelaskan di atas;
3.3.4 **Penerapan algoritma**, yaitu pelatihan dan pengujian model *Naive Bayes* dan SVM pada dataset yang telah diproses;
3.3.5 **Evaluasi hasil**, berupa perbandingan kinerja kedua algoritma menggunakan confusion matrix; dan
3.3.6 **Penyusunan hasil serta penarikan kesimpulan**, berdasarkan temuan perbandingan akurasi, presisi, dan recall dari kedua metode.

4. PEMBAHASAN

Penelitian ini berhasil mengumpulkan 81 data tanggapan mahasiswa peserta Program Kampus Mengajar yang dikumpulkan pada tanggal 15 Mei 2025 melalui kuesioner. Data tersebut kemudian diberi label secara manual sebagai sentimen positif atau negatif, sebelum diproses melalui tahapan preprocessing dan diklasifikasikan menggunakan algoritma *Naive Bayes* (NB) dan Support Vector Machine (SVM) dengan rasio data latih dan data uji sebesar 70:30.

Tabel 4. Data Hasil Klasifikasi Naïve Bayes dan SVM

No	Kategori Sentimen	Jumlah Data	Persentase
1	Positif	61	75,3%
2	Negatif	20	24,7%
Total		**81**	**100%**

Mayoritas mahasiswa memberikan tanggapan positif terhadap program ini, terutama terkait peningkatan soft skill seperti kerja sama tim, kepemimpinan, dan komunikasi, serta pengalaman nyata menghadapi dinamika dunia pendidikan secara langsung. Meski demikian, terdapat 20 tanggapan negatif yang didominasi keluhan mengenai beban administrasi yang tinggi, dosen

pembimbing lapangan yang kurang aktif memberi arahan, keterbatasan fasilitas di lokasi penempatan, serta koordinasi program yang dinilai belum maksimal.

Perbandingan Hasil Klasifikasi *Naive Bayes* dan SVM Setelah data diklasifikasikan menggunakan kedua algoritma pada data uji sebanyak 36 data (30% dari total data), diperoleh hasil confusion matrix sebagai berikut:

Tabel 5. Hasil Perbandingan Klasifikasi Naïve Bayes dan SVM

Matrik Evaluasi	Naïve Bayes	Support Vector Machine
Akurasi	83,33%	86,11%
Presisi (Positif)	77,7%	100%
Presisi (Negatif)	88,9%	72,22%
Recall (Positif)	87,5%	78,26%
Recall (Negatif)	80%	100%

Dari tabel di atas terlihat bahwa SVM unggul dalam akurasi keseluruhan (86,11% berbanding 83,33%), yang berarti SVM secara umum lebih tepat dalam mengklasifikasikan data uji dibandingkan *Naive Bayes.* Namun, jika ditelusuri lebih dalam pada metrik presisi dan recall, *Naive Bayes* menunjukkan performa yang lebih seimbang, khususnya dalam mendeteksi sentimen positif (presisi 77,7% dan recall 87,5%, dengan selisih hanya sekitar 10 poin). Sebaliknya, SVM mencapai presisi sempurna 100% untuk sentimen positif, tetapi recall-nya hanya 78,26% — artinya SVM sangat jarang salah ketika memprediksi data sebagai positif, namun cenderung melewatkan sejumlah data positif yang sebenarnya (menganggapnya sebagai negatif). Pola serupa juga terlihat pada sentimen negatif, di mana SVM memiliki recall sempurna 100% tetapi presisi lebih rendah (72,22%), yang berarti SVM sering salah mengklasifikasikan data positif sebagai negatif.

Temuan ini mengindikasikan bahwa meskipun SVM lebih unggul secara akurasi agregat, *Naive Bayes* memiliki keunggulan tersendiri dalam konsistensi dan keseimbangan performa klasifikasi antar kelas, sehingga pemilihan algoritma "terbaik" tidak bisa hanya didasarkan pada satu metrik akurasi saja, melainkan perlu mempertimbangkan kebutuhan spesifik penelitian — apakah lebih mengutamakan ketepatan prediksi positif (presisi tinggi) atau kemampuan menangkap seluruh data positif (recall tinggi).

Perbandingan dengan Penelitian Terdahulu untuk memperkuat validitas temuan, hasil penelitian ini dibandingkan dengan beberapa studi terdahulu yang juga membandingkan algoritma NB dan SVM pada domain analisis sentimen berbeda:

Tabel 6.Perbandingan dengan Penelitian Terdahulu

No	Peneliti (Tahun)	Topik	Akurasi NB	Akurasi SVM
1	Saputra et al., (2023)	Sentimen Piala Dunia FIFA 2022 (Twitter)	82%	85%
2	Matheos Sarimole et al., (2024)	Sentimen Aplikasi Satu Sehat (Twitter)	81,65%	87,95%
3	Setiawan et al., (2024)	Analisis Sentimen (Twitter)	77%	84%
4	**Penelitian ini (2025)**	**Sentimen Program Kampus Mengajar**	**83,33%**	**86,11%**

Ketiga penelitian pembanding di atas secara konsisten menunjukkan bahwa SVM mengungguli *Naive Bayes* dalam hal akurasi, sejalan dengan hasil yang diperoleh dalam penelitian ini. Konsistensi pola ini memperkuat argumen bahwa SVM cenderung lebih unggul dalam menangani klasifikasi teks berbasis opini secara umum, meskipun besaran selisih akurasi antara kedua metode dapat bervariasi tergantung karakteristik dataset, jumlah data, dan sumber data (media sosial vs. kuesioner langsung seperti pada penelitian ini).

Dampak Penelitian Penelitian ini memberikan sejumlah dampak yang signifikan, baik secara akademik maupun praktis:

1. Dampak Akademik/Ilmiah Penelitian ini memperkaya literatur analisis sentimen berbasis machine learning, khususnya pada domain evaluasi program pendidikan yang selama ini lebih banyak diteliti pada konteks media sosial seperti Twitter. Temuan bahwa SVM unggul dalam akurasi namun *Naive Bayes* lebih stabil dalam presisi-recall, memberikan panduan metodologis bagi peneliti selanjutnya dalam memilih algoritma klasifikasi yang tepat sesuai karakteristik dataset dan tujuan analisis bukan sekadar mengejar akurasi tertinggi, tetapi juga mempertimbangkan keseimbangan performa antar kelas.
2. Dampak Praktis bagi Penyelenggara Program Hasil analisis sentimen dalam penelitian ini dapat langsung dimanfaatkan oleh pihak penyelenggara Program Kampus Mengajar sebagai bahan evaluasi berbasis data. Proporsi tanggapan positif yang tinggi (75,3%) menjadi indikator keberhasilan program dalam memberikan pengalaman belajar bermakna, sementara tanggapan negatif (24,7%) terutama terkait bimbingan lapangan dan beban administrasi menjadi masukan konkret untuk perbaikan pelaksanaan program di periode berikutnya.
3. Dampak Metodologis Penelitian ini membuktikan bahwa pendekatan berbasis survei digital (kuesioner) yang dikombinasikan dengan teknik machine learning mampu menghasilkan evaluasi program pendidikan yang lebih objektif, efisien, dan terukur dibandingkan metode evaluasi konvensional yang bersifat kualitatif-manual. Pendekatan ini membuka peluang bagi institusi pendidikan atau organisasi lain untuk menerapkan metode serupa dalam mengevaluasi program-program berbasis opini publik secara lebih sistematis dan berkelanjutan.

5. KESIMPULAN

Penelitian ini bertujuan untuk mengidentifikasi kecenderungan sentimen mahasiswa terhadap Program Kampus Mengajar sekaligus membandingkan kinerja dua algoritma klasifikasi teks, yaitu *Naive Bayes (NB)* dan *Support Vector Machine (SVM)*, dalam mengklasifikasikan opini tersebut ke dalam kategori positif dan negatif. Berdasarkan hasil pengumpulan data melalui kuesioner terhadap 81 mahasiswa peserta program, serta proses analisis yang telah dilakukan mulai dari preprocessing hingga evaluasi model, penelitian ini menghasilkan beberapa kesimpulan sebagaimana diuraikan berikut.

Pertama, dari segi persepsi mahasiswa, mayoritas responden memberikan tanggapan positif terhadap Program Kampus Mengajar. Dari total 81 data yang diperoleh, sebanyak 61 tanggapan (75,3%) bersifat positif dan 20 tanggapan (24,7%) bersifat negatif. Tanggapan positif umumnya berkaitan dengan pengembangan soft skill seperti kerja sama tim, kepemimpinan, dan komunikasi, serta pengalaman nyata dalam menghadapi dunia pendidikan secara langsung. Sementara itu, tanggapan negatif lebih banyak menyoroti beban administrasi yang tinggi, minimnya bimbingan dari dosen pembimbing lapangan, serta koordinasi program yang dinilai belum optimal. Temuan ini menunjukkan bahwa meskipun program ini secara umum dinilai berhasil oleh mahasiswa, masih terdapat aspek teknis pelaksanaan yang perlu dievaluasi dan diperbaiki.

Kedua, dari segi perbandingan algoritma, hasil pengujian klasifikasi menunjukkan bahwa algoritma SVM memperoleh akurasi lebih tinggi sebesar 86,11% dibandingkan *Naive Bayes* yang mencapai akurasi 83,33%. Meskipun demikian, jika ditinjau dari metrik presisi dan recall, *Naive Bayes* menunjukkan performa yang lebih seimbang, khususnya dalam mendeteksi sentimen positif dengan presisi 77,7% dan recall 87,5%. Sebaliknya, SVM meskipun mencapai presisi sempurna 100% untuk sentimen positif, memiliki recall yang lebih rendah yaitu 78,26%, yang berarti model ini cenderung melewatkan sejumlah data positif yang sebenarnya. Hal ini menegaskan bahwa keunggulan suatu algoritma tidak dapat dinilai hanya dari satu metrik akurasi semata, melainkan perlu mempertimbangkan keseimbangan performa antar kelas sentimen secara menyeluruh.

Ketiga, berdasarkan perbedaan hasil akurasi, presisi, dan recall yang jelas antara kedua metode tersebut, dapat disimpulkan bahwa hipotesis alternatif (H_1) dalam penelitian ini diterima, yaitu terdapat perbedaan yang signifikan pada hasil klasifikasi sentimen antara algoritma *Naive Bayes* dan *Support Vector Machine* ketika diterapkan pada data opini mahasiswa terhadap Program Kampus Mengajar. Dengan demikian, hipotesis nol (H_0) yang menyatakan tidak adanya perbedaan berarti antara kedua metode tersebut ditolak. Temuan ini juga konsisten dengan sejumlah penelitian terdahulu yang membandingkan NB dan SVM pada domain analisis sentimen berbeda, yang secara umum menunjukkan bahwa SVM cenderung unggul dalam hal akurasi.

Secara keseluruhan, penelitian ini berhasil menjawab rumusan masalah yang diajukan, yaitu mengungkap kecenderungan sentimen mahasiswa yang didominasi tanggapan positif, sekaligus membuktikan bahwa pemilihan algoritma klasifikasi berpengaruh signifikan terhadap hasil analisis sentimen. Temuan ini diharapkan dapat menjadi acuan bagi penyelenggara Program Kampus Mengajar dalam melakukan evaluasi dan perbaikan program secara berkelanjutan, serta menjadi referensi metodologis bagi peneliti selanjutnya dalam memilih algoritma klasifikasi teks yang tepat sesuai karakteristik data dan tujuan penelitian, termasuk kemungkinan penggunaan dataset yang lebih besar dan beragam serta perbandingan dengan algoritma klasifikasi lainnya di masa mendatang.

DAFTAR PUSTAKA

[1] Anwar, K. (2022). Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *2*(4), 148–155. https://doi.org/10.30865/klik.v2i4.315

[2] Anwar, R. N. (2021). Pelaksanaan Kampus Mengajar Angkatan 1 Program Merdeka Belajar Kampus Merdeka di Sekolah Dasar. *Jurnal Pendidikan Dan Kewirausahaan*, *9*(1), 210–219. https://doi.org/10.47668/pkwu.v9i1.221

[3] Apif Supriadi, & Fatmasari. (2021). Implementasi Metode Klasifikasi Naive Bayes Pada Sistem Analisis Opini Pengguna Twitter Berbasis Web. *Jurnal Sistem Informasi*, *10*(1), 46–54. https://doi.org/10.51998/jsi.v10i1.356

[4] Desiani, A., Zayanti, D. A., Ramayanti, I., Ramadhan, F. F., Sriwijaya, U., Kedokteran, I., Kedokteran, F., & Palembang, U. M. (2025). *PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN COMPARISON OF SUPPORT VECTOR MACHINE (SVM) AND LOGISTIC*. *4*(1).

[5] Firdaus, M. R., Rizki, F. M., Gaus, F. M., & Susanto, I. K. (2020). Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *4*(1), 66. https://doi.org/10.30645/j-sakti.v4i1.188

[6] Indrayuni, E. (2018). Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film. *Jurnal Pilar Nusa Mandiri*, *14*(2), 175. https://doi.org/10.33480/pilar.v14i2.918

[7] Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization.

Jurnal Nasional Teknik Elektro Dan Teknologi Informasi, *9*(2), 162–170. https://doi.org/10.22146/jnteti.v9i2.102

[8] Ramadinah, S. (2021). *Perbandingan Metode Klasifikasi Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Review Gojek*.

[9] Rina Noviana, & Isram Rasal. (2023). Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter. *Jurnal Teknik Dan Science*, *2*(2), 51–60. https://doi.org/10.56127/jts.v2i2.791

[10] Rozaq, A., Yunitasari, Y., Sussolaikah, K., & Sari, E. R. N. (2022). Sentiment Analysis of Kampus Mengajar 2 Toward the Implementation of Merdeka Belajar Kampus Merdeka Using Naïve Bayes and Euclidean Distence Methods. *International Journal of Advances in Data and Information Systems*, *3*(1), 30–37. https://doi.org/10.25008/ijadis.v3i1.1233

[11] Salim, E., & Syafrullah, M. (2023). *Jakarta Barat Menggunakan Algoritme K-Nearest Neighbor*. *20*(1), 58–65. https://kemsalim.space/ulasan_dukcapil/

[12] Sasmita, A. B., Rahayudi, B., & Muflikhah, L. (2022). Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(3), 1208–1214. http://j-ptiik.ub.ac.id

[13] Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 183–192. https://doi.org/10.29408/edumatic.v8i1.25667

[14] Simatupang, E., & Yuhertiana, I. (2021). Merdeka Belajar Kampus Merdeka terhadap Perubahan Paradigma Pembelajaran pada Pendidikan Tinggi: Sebuah Tinjauan Literatur. *Jurnal Bisnis, Manajemen, Dan Ekonomi*, *2*(2), 30–38. https://doi.org/10.47747/jbme.v2i2.230

[15] Syafii Imam Muhamad. (2023). Sentimen Analisis Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (Nbc). *Teknologipintar.Org*, *3*(2), 1.

[16] Wardani, S. (2015). Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan Svm. *Jurnal Dinamika Informatika*, *5*(November), 1–13.

[17] Widowati, T. T., & Sadikin, M. (2021). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, *11*(2), 626–636. https://doi.org/10.24176/simet.v11i2.4568

Biodata Penulis

Nama Lengkap Penulis Pertama, dapat berisikan tempat dan / atau tanggal lahir. Berikutnya narasi mengenai latar belakang pendidikan penulis yang mencakup jeis gelar dalam bidang apa, lembaga, kota, negara, dan tahun gelar itu diperoleh. Selanjutnya yang perlu dicantumkan adalah pekerjaan/kesibukan penulis saat ini. Jika disediakan sebuah foto, maka biografi di sekitarnya akan menjorok dan foto ditempatkan di kiri atas dari biografi.

Nama Lengkap Penulis Kedua, dapat berisikan tempat dan / atau tanggal lahir. Berikutnya narasi mengenai latar belakang pendidikan penulis yang mencakup jeis gelar dalam bidang apa, lembaga, kota, negara, dan tahun gelar itu diperoleh. Selanjutnya yang perlu dicantumkan adalah pekerjaan/kesibukan penulis saat ini. Jika disediakan sebuah foto, maka biografi di sekitarnya akan menjorok dan foto ditempatkan di kiri atas dari biografi.

Nama Lengkap Penulis Ketiga, dapat berisikan tempat dan / atau tanggal lahir. Berikutnya narasi mengenai latar belakang pendidikan penulis yang mencakup jeis gelar dalam bidang apa, lembaga, kota, negara, dan tahun gelar itu diperoleh. Selanjutnya yang perlu dicantumkan adalah

pekerjaan/kesibukan penulis saat ini. Jika disediakan sebuah foto, maka biografi di sekitarnya akan menjorok dan foto ditempatkan di kiri atas dari biografi.