

Analisis Sentimen Persepsi Pengguna Kurikulum Merdeka Belajar menggunakan Metode *Naive Bayes Classifier* (NBC) Dan *Support Vector Machine* (SMV)

Tia Millenia Alvionita¹, Achmad Hindasyah^{1,2}, Abu Khalid^{1,3}

¹Program Studi Magister Teknik Informatika, Program Pascasarjana, Universitas Pamulang

²Badan Riset Dan Inovasi Nasional (BRIN)

tyamilleniaalvionita@gmail.com,

Diterima : 30 Agustus 2025

Disetujui : 20 September 2025

Abstract— Penerapan Kurikulum Merdeka di berbagai jenjang pendidikan menimbulkan beragam persepsi dari para penggunanya, termasuk di tingkat Sekolah Menengah Kejuruan (SMK). Perubahan pendekatan pembelajaran, integrasi proyek penguatan profil pelajar Pancasila, serta fleksibilitas dalam perancangan kurikulum menjadi faktor-faktor yang memicu variasi tanggapan dari guru dan siswa sebagai pelaksana langsung. Penelitian ini bertujuan untuk mengevaluasi persepsi tersebut melalui analisis sentimen berbasis teks, dengan tujuan mengetahui kecenderungan opini pengguna terhadap Kurikulum Merdeka di SMK Teknikom Cikarang. Pendekatan yang digunakan adalah kuantitatif dengan metode klasifikasi teks menggunakan Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM). Data diperoleh dari tanggapan siswa melalui survei terbuka, yang kemudian dikategorikan ke dalam tiga label sentimen: positif, negatif, dan netral. Penelitian ini juga mengidentifikasi faktor-faktor utama yang memengaruhi persepsi siswa, di antaranya: pemahaman terhadap alur tujuan pembelajaran (ATP), pengalaman belajar berbasis proyek, keterlibatan guru dalam memfasilitasi pembelajaran, serta kemudahan dalam penggunaan media ajar digital. Metode klasifikasi diuji menggunakan tiga varian NBC, yaitu GaussianNB, MultinomialNB, dan BernoulliNB, kemudian dibandingkan dengan algoritma SVM. Hasil evaluasi berdasarkan confusion matrix menunjukkan bahwa metode SVM memiliki akurasi tertinggi sebesar 96,5%, dengan klasifikasi benar sebanyak 36 data positif, 45 data negatif, dan 112 data netral. Di antara model Naive Bayes, performa terbaik ditunjukkan oleh BernoulliNB dengan akurasi 92,5%, diikuti oleh MultinomialNB dengan akurasi yang sama, sementara GaussianNB menunjukkan performa terendah dengan akurasi 77,5%. Keunggulan SVM disebabkan oleh kemampuannya dalam menangani data teks berdimensi tinggi dan bersifat sparsity, serta ketahanannya dalam membedakan kelas sentimen yang memiliki kemiripan—terutama ketika fitur teks tidak mengikuti distribusi probabilistik yang kuat seperti yang diasumsikan dalam model Naive Bayes. Temuan ini menunjukkan bahwa Support Vector Machine lebih unggul dalam mengklasifikasikan persepsi pengguna terhadap Kurikulum Merdeka Belajar dibandingkan dengan varian Naive Bayes. Hasil ini diharapkan dapat menjadi dasar dalam pengembangan sistem evaluasi berbasis analisis sentimen di bidang pendidikan, khususnya untuk menilai implementasi kurikulum baru berdasarkan opini siswa, kemampuan generalisasi tinggi SVM mampu menemukan hyperplane optimal yang memisahkan data dengan margin maksimum, sehingga meminimalkan kesalahan klasifikasi.

Keywords — Analisis Sentimen, Kurikulum Merdeka Belajar, Naive Bayes Classifier (NBC), Support Vector Machine (SVM), SMK Teknikom Cikarang

I. PENDAHULUAN

Kurikulum Merdeka Belajar adalah sebuah inisiatif baru dalam sistem pendidikan di

Indonesia yang bertujuan untuk memberikan lebih banyak kebebasan dan fleksibilitas kepada sekolah dan guru dalam mengelola

proses pembelajaran. Kurikulum ini diperkenalkan sebagai respon terhadap kebutuhan akan pendidikan yang lebih adaptif dan relevan dengan perkembangan zaman, serta untuk mendorong pembelajaran yang lebih kontekstual dan bermakna bagi siswa. Namun, seperti halnya setiap perubahan besar dalam kebijakan pendidikan, implementasi Kurikulum Merdeka telah memicu berbagai reaksi dan persepsi dari guru dan siswa, baik positif, netral maupun negatif. [1]. Penting untuk memahami persepsi Siswa terhadap Kurikulum Merdeka Belajar agar pembuat kebijakan dapat mengevaluasi dan menyesuaikan implementasi kurikulum tersebut sesuai dengan kebutuhan dan harapan publik. Analisis sentimen menawarkan cara yang efektif untuk mengukur dan mengkategorikan opini publik yang diekspresikan melalui berbagai platform, seperti media sosial, forum online, dan komentar berita [2].

Dalam konteks ini, metode pembelajaran mesin seperti Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM) dapat digunakan untuk melakukan analisis sentimen secara otomatis dan efisien. SVM adalah metode yang dikenal efektif dalam menangani masalah klasifikasi dengan dimensi tinggi dan sering memberikan hasil yang baik pada tugas klasifikasi teks. Sementara itu, NBC adalah metode probabilistik yang sederhana namun efektif, yang sering digunakan dalam tugas klasifikasi teks karena kemampuannya untuk menangani dataset yang besar dengan komputasi yang relatif ringan. [3].

Penelitian ini bertujuan untuk mengaplikasikan kedua metode tersebut dalam analisis sentimen persepsi Siswa terhadap Kurikulum Merdeka. Dengan membandingkan kinerja NBC dan SVM, penelitian ini tidak hanya memberikan wawasan tentang sentimen umum siswa tetapi juga menilai keefektifan masing-masing metode dalam konteks ini. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang berharga bagi pembuat kebijakan dan pemangku kepentingan pendidikan dalam mengevaluasi dan mengoptimalkan implementasi Kurikulum

Merdeka. [4]. Analisis sentimen adalah proses untuk mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam sebuah teks. Analisis ini sering digunakan untuk memahami persepsi atau pandangan pengguna terhadap suatu produk, layanan, atau kebijakan tertentu, seperti Kurikulum Merdeka dalam kasus ini. Metode yang sering digunakan untuk analisis sentimen adalah Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM). [5]

II. TINJAUAN LITERATURE

Sebelum memulai penelitian ini, kajian terhadap penelitian sebelumnya perlu dilakukan agar dapat mengetahui metode, data, dan model yang sudah pernah digunakan. Tinjauan pustaka ini menjadi landasan penelitian untuk memahami permasalahan dalam penerapan metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM) dalam analisis sentimen persepsi pengguna terhadap Kurikulum Merdeka di SMK Teknikom Cikarang [6]. Berbagai penelitian telah dilakukan terkait penerapan metode NBC, penelitian penerapan metode SVM, serta penelitian terkait analisis sentimen atau prediksi berdasarkan sentimen pengguna. Di bawah ini ditampilkan beberapa penelitian yang relevan dalam bentuk tabel untuk memberikan gambaran menyeluruh mengenai studi-studi yang relevan, teknik yang diterapkan, dan temuan-temuan utama yang dapat mendukung penelitian ini. [7]. Naive Bayes adalah suatu klasifikasi berpeluang yang berdasarkan aplikasi teorema Bayes. Naïve Bayes merupakan algoritma klasifikasi yang efektif dalam mendapatkan hasil yang akurat dan efisien dalam proses penalaran memanfaatkan input yang ada dengan cara yang relatif cepat. Algoritma ini bertujuan untuk melakukan klasifikasi data pada kelas tertentu (Wibowo, 2020). [8].

Dalam penerapannya metode naïve bayes berhubungan dengan machine learning. Naïve Bayes adalah metode klasifikasi atau teknik machine learning yang populer/umum digunakan dalam klasifikasi teks, memiliki performa yang baik pada banyak domain,

sederhana dan efisien. Namun naïve bayes sangat sensitif dalam pemilihan fitur, terlalu banyak jumlah fitur dapat mengakibatkan meningkatnya waktu perhitungan dan menurunkan akurasi klasifikasi [9]. Support Vector Machine (SVM) adalah teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Handayani, et al., 2020). SVM termasuk dalam jenis klasifikator yang biner, linier, dan non probabilistik. Pada umumnya hal pertama yang mendasari untuk memahami klasifikasi dengan SVM adalah mencari garis (hyperplane) yang optimal (Mutawalli, et al., 2019). Hal ini berfungsi untuk memisahkan dua kelas data yang berbeda, yaitu positif (+1) dan negatif (-1). Pada gambar 2.4 untuk data positif (+1) disimbolkan dengan warna kuning dan data negatif (-1) disimbolkan dengan warna merah. Secara umum, proses SVM diilustrasikan dalam gambar 2.4. Grafik sebelah kiri pada gambar 2.4 mengilustrasikan tentang beberapa kemungkinan garis pemisah (discrimination boundaries) pada SVM untuk melakukan set data.

Sedangkan pada grafik sebelah kanan mengilustrasikan discrimination boundaries dengan margin maksimum. Margin atau juga yang disebut juga batas pemisah adalah jarak antara dua kelas data terdekat pada bidang hyperplane. Hyperplane dengan margin terbaik yang akan memberikan generalisasi untuk mendapatkan hasil klasifikasi yang lebih baik.

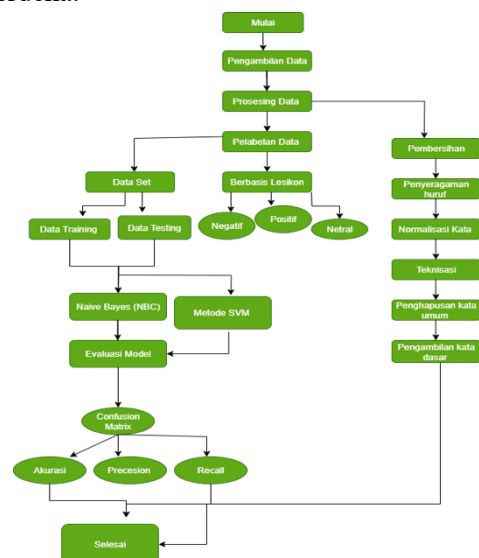
Untuk memperkuat hasil klasifikasi sentimen menggunakan metode Naive Bayes, analisis ini juga menggunakan metode Support Vector Machine (SVM) untuk meningkatkan akurasi dalam mengklasifikasikan sentimen siswa terhadap Kurikulum Merdeka. SVM akan mengidentifikasi dan memisahkan data pada kelas-kelas sentimen yang berbeda, misalnya "positif," "negatif," atau "netral," dengan menemukan garis pemisah terbaik (hyperplane) yang memaksimalkan margin antar-kelas[10].

Proses SVM ini dilakukan dengan cara memilih data sentimen dari survei siswa SMK Teknikom Cikarang dan menentukan pola yang optimal untuk memisahkan masing-masing

kelas sentimen. Setelah proses klasifikasi selesai, hasil dari SVM akan digunakan sebagai model pembanding untuk hasil Naive Bayes, sehingga bisa dipastikan bahwa hasil analisis sentimen benar-benar mencerminkan persepsi siswa dengan akurasi yang lebih baik. Berdasarkan pendekatan sistem pendukung keputusan, analisis ini menggunakan struktur model berbasis data serta metode komputasi SVM yang memberikan beberapa alternatif untuk mendukung keputusan mengenai persepsi siswa[10].

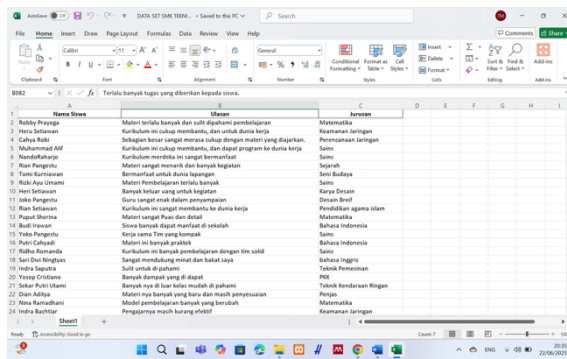
III. METODE PENELITIAN

Penelitian ini akan menerapkan metode SVM dan NBC dalam menganalisis sentimen persepsi pengguna terhadap penerapan Kurikulum Merdeka di SMK Teknikom Cikarang. Metode Naive Bayes digunakan untuk mengklasifikasikan sentimen pengguna ke dalam kategori "positif" atau "negatif." Selanjutnya, untuk memperkuat hasil klasifikasi metode Naive Bayes, metode Support Vector Machine (SVM) digunakan untuk memperkirakan tingkat akurasi sentimen dan memberikan penguatan pada hasil klasifikasi secara keseluruhan. Dengan demikian, penelitian ini akan melakukan analisis kebutuhan dan spesifikasi teknis yang diperlukan dalam menyelesaikan masalah terkait sentimen pengguna terhadap Kurikulum Merdeka.



Gambar 1. Flowchart Analisis Sentimen Merdeka Belajar

B. Pengumpulan Data dan Persiapan Data



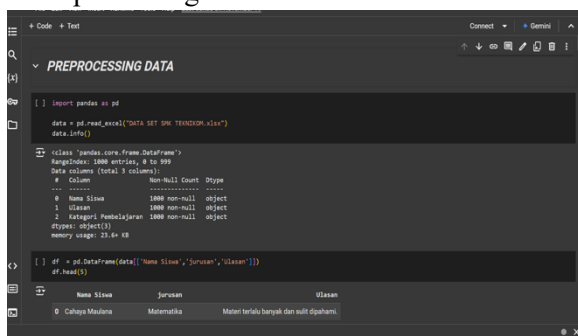
	Nama Siswa	Jurusan	Ulasan
1			
2	Reza Prasetya	Matematika	Materi terlalu banyak dan sulit dipahami pembelajaran
3	Nisa Setiawan	Kejuruan	Kurikulum ini cukup membantu, dan untuk dunia kerja
4	Calvin Rabi	Kejuruan	Sebagian besar sangat sesuai dengan materi yang diajarkan
5	Mahmud Alif	Kejuruan	Kurikulum ini cukup membantu, dan dapat diterapkan ke dunia kerja
6	Handika Raka	Kejuruan	Kurikulum merdeka ini sangat bermanfaat
7	Rian Pangestika	Kejuruan	Materi sangat menarik dan banyak kegiatan
8	Toni Kurniawan	Kejuruan	Bermanfaat untuk dunia lapangan
9	Kiki Nur Izzah	Kejuruan	Materi pembelajaran terlihat banyak
10	Nani Setiawan	Kejuruan	Banyak belajar yang untuk kegiatan
11	Adha Pangestika	Kejuruan	Guru sangat baik dalam penyampaian
12	Rian Setiawan	Kejuruan	Kurikulum ini sangat membantu ke dunia kerja
13	Paput Setiawan	Kejuruan	Materi sangat bagus dan detail
14	Budi Irawan	Kejuruan	Siswa banyak dapat manfaat di sekolah
15	Yoko Pangestika	Kejuruan	Kepa sama Tim yang banyak
16	Putri Cahyadi	Kejuruan	Materi ini banyak praktik
17	Rafika Ramadani	Kejuruan	Kurikulum ini banyak pembelajaran dengan tim solid
18	Sari Nur Hafidha	Kejuruan	Sangat mendukung minat dan bakat siswa
19	Indira Saputra	Kejuruan	Suatu untuk di pahami
20	Yenny Cicihama	Kejuruan	Banyak kegiatan yang di dapat
21	Putri Nur Izzah	Kejuruan	Banyak nya di saat belajar mudah di pahami
22	Dian Adhika	Kejuruan	Materi nya banyak yang baru dan mudah dipahami
23	Nisa Ramadhani	Kejuruan	Model pembelajaran banyak yang berubah
24	Indira Bahtiar	Kejuruan	Pengajaran masih kurang efektif

Gambar 2. Dataset

Data penelitian yang digunakan adalah dataset SMK TEKNIKOM CIKARANG Tahun ajaran 2024/2025. Berikut tampilan dataset yang digunakan. Adapun dataset lengkap terdapat dalam lampiran. Tidak semua atribut dataset penelitian digunakan. Hanya atribut yang relevan terhadap penelitian saja yang digunakan. Atribut yang relevan terhadap penelitian diperoleh dari hasil wawancara dengan siswa SMK TEKNIKOM CIKARANG. Berikut atribut yang digunakan dalam penelitian.

Selanjutnya agar dataset dapat diproses dengan cepat pada python, maka file dataset dengan ekstensi file .xls akan diubah dalam bentuk .csv. Proses perubahan ekstensi file dari .xls ke .csv menggunakan Microsoft Excel dengan hasil kapasitas file menjadi sangat kecil, dari berukuran 84 KB menjadi 28 KB.

C. Preprocessing Data



```
[ ] Import pandas as pd
data = pd.read_excel("DATA SET SMK TEKNIKOM CIKARANG.xlsx")
data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1000 entries, 0 to 999
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   Nama Siswa  1000 non-null    object
 1   Jurusan    1000 non-null    object
 2   Ulasan     1000 non-null    object
dtypes: object(3)
memory usage: 23.4+ KB

[ ] df = pd.DataFrame(data[['Nama Siswa', 'Jurusan', 'Ulasan']])
df.head(5)
```

	Nama Siswa	Jurusan	Ulasan
0	Calvin Rabi	Matematika	Materi terlalu banyak dan sulit dipahami

Gambar 3. Hasil Dataset Yang Sudah di cleaning

Tampak pada gambar di atas, terdapat library yang diimport, yaitu library pandas yang selanjutnya dipanggil sebagai "pd". Library ini berfungsi untuk membaca, mengolah, dan

menganalisis data. Pada gambar tersebut, dataset bernama "DATA SET SMK TEKNIKOM.xlsx" dibaca menggunakan perintah `pd.read_excel()` dan disimpan dalam sebuah data frame bernama `data`.

Selanjutnya, digunakan perintah `data.info()` untuk mengetahui informasi terkait dataset yang digunakan dalam penelitian. Informasi tersebut meliputi jumlah entri, jumlah kolom, nama kolom, tipe data di setiap kolom, dan jumlah data non-null dalam setiap kolom. Selain itu, terdapat proses membuat data frame baru bernama `df` dengan menyaring kolom "Nama Siswa", "Jurusan", dan "Ulasan". Hasil dari data frame ini ditampilkan menggunakan perintah `df.head(5)` yang menampilkan 5 baris pertama dari data tersebut.

1. Data duplikat

Gambar tersebut menunjukkan hasil dari perintah `df.info()` yang digunakan untuk menampilkan informasi umum mengenai dataset yang digunakan dalam penelitian ini. Berdasarkan output tersebut, diketahui bahwa dataset terdiri dari 1.000 baris data dengan tiga kolom, yaitu Nama Siswa, Ulasan, dan Jurusan. Seluruh kolom memiliki nilai yang lengkap (non-null), yang berarti tidak terdapat data kosong di dalamnya [13]

2. Wordcloud Sebelum Data Prosesing

ada tahap awal analisis sentimen terhadap persepsi siswa mengenai Kurikulum Merdeka, dilakukan visualisasi data dalam bentuk WordCloud. Visualisasi ini bertujuan untuk melihat kata-kata yang paling sering muncul dalam kumpulan ulasan siswa sebelum dilakukan proses preprocessing. Data yang digunakan berasal dari kolom "Ulasan" yang memuat komentar dan pendapat siswa terhadap implementasi Kurikulum Merdeka di lingkungan SMK Teknikom Cikarang. Setiap ulasan dikumpulkan menjadi satu teks panjang, kemudian divisualisasikan dalam bentuk WordCloud [14]

3. Data Cleaning

Pada tahap pra-pemrosesan data teks, salah satu langkah penting adalah penghapusan simbol-simbol atau karakter non-alfanumerik yang tidak memberikan kontribusi bermakna dalam

analisis sentimen. Simbol seperti tanda baca, karakter khusus, dan simbol non-standar sering kali muncul dalam teks ulasan, tetapi umumnya tidak relevan untuk diproses oleh model klasifikasi sentimen berbasis machine learning. Meningkatkan akurasi analisis sentimen, dengan mengurangi gangguan dari simbol-simbol yang tidak diperlukan.

Menstandarkan format teks sehingga lebih mudah untuk diproses pada tahap berikutnya, seperti tokenisasi, stopwords removal, dan stemming.

Mengoptimalkan performa model klasifikasi, khususnya dalam pendekatan berbasis machine learning seperti Naive Bayes dan Support Vector [15].

4. Normalisasi Kata

Dalam analisis teks, terutama pada data ulasan yang ditulis secara bebas oleh siswa, sering ditemukan penggunaan kata-kata tidak baku, bahasa gaul, singkatan, atau kesalahan penulisan. Hal ini dapat menurunkan akurasi model analisis sentimen, karena algoritma akan menganggap kata tidak baku dan kata baku sebagai dua entitas yang berbeda. Oleh karena itu, dilakukan proses normalisasi, yaitu mengubah kata tidak baku menjadi bentuk baku sesuai kaidah bahasa Indonesia [16].

5. Tokenization

Setelah proses normalisasi dilakukan, tahapan selanjutnya dalam pra-pemrosesan data teks adalah tokenisasi. Tokenisasi merupakan proses memecah teks atau kalimat menjadi satuan kata-kata kecil yang disebut sebagai token. Setiap token umumnya merepresentasikan satu kata yang akan dianalisis lebih lanjut oleh model analisis sentimen [17].

6. Proses Stopword Removal

Setelah melalui proses tokenisasi, tahapan selanjutnya dalam pra-pemrosesan data ulasan siswa terhadap Kurikulum Merdeka adalah penghapusan stopwords (stopword removal). Stopword adalah kata-kata umum dalam bahasa yang tidak memiliki makna signifikan dalam konteks analisis, seperti: yang, dan, di, ke, dari, itu, dan sebagainya.

Meskipun kata-kata ini penting secara tata bahasa, namun dalam konteks analisis sentimen

atau penggalan opini, stopwords justru dapat mengganggu proses ekstraksi makna utama dari ulasan. Oleh karena itu, kata-kata tersebut perlu dihapus agar analisis fokus pada kata-kata bermakna atau emosional yang mencerminkan persepsi siswa

7. Steaming Data

Setelah dilakukan tahapan tokenisasi dan penghapusan stopwords, proses selanjutnya dalam pra-pemrosesan data ulasan siswa terhadap Kurikulum Merdeka adalah stemming, atau dalam Bahasa Indonesia dikenal dengan istilah peluruhan kata.

Stemming adalah proses mengubah kata turunan atau kata berimbuhan menjadi bentuk dasar (root word). Contohnya, kata-kata seperti "pembelajaran", "belajarannya", dan "mempelajari" akan dipetakan menjadi bentuk dasar "belajar". Hal ini dilakukan untuk menyamakan makna kata-kata yang berbeda bentuk tetapi memiliki akar yang sama.

8. Wordcloud Setelah Proseing

Setelah melalui serangkaian proses preprocessing seperti pembersihan simbol, case folding, normalisasi kata tidak baku, tokenisasi, penghapusan stopwords, dan stemming, dilakukan visualisasi data ulasan siswa dalam bentuk Wordcloud. Wordcloud merupakan salah satu metode visualisasi teks yang menampilkan kumpulan kata dengan ukuran huruf yang proporsional terhadap frekuensi kemunculannya. Semakin besar ukuran suatu kata dalam Wordcloud, maka semakin sering kata tersebut muncul dalam data ulasan.

Tujuan dibuatnya Wordcloud setelah tahap preprocessing antara lain:

- Untuk menggambarkan kata-kata kunci yang paling dominan muncul dalam ulasan siswa terhadap Kurikulum Merdeka.
- Untuk memberikan gambaran visual yang intuitif terhadap distribusi kata yang mengandung opini atau persepsi siswa.
- Sebagai dasar observasi awal dalam menganalisis arah sentimen siswa sebelum dilakukan klasifikasi menggunakan metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM)

D. Data Labelling Data

Dalam analisis sentimen, pelabelan data merupakan tahapan penting yang dilakukan sebelum proses pelatihan model klasifikasi. Tujuan dari pelabelan data adalah untuk memberikan kategori sentimen terhadap setiap data teks (ulasan) berdasarkan isi atau maknanya. Pelabelan ini akan menjadi variabel target (label) dalam proses pembelajaran mesin yang digunakan oleh algoritma klasifikasi seperti Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM). Kategori Label Sentimen,

Sentimen Positif: Ulasan yang mengandung kata-kata apresiasi, pengalaman menyenangkan, atau dukungan terhadap Kurikulum Merdeka.

Sentimen Negatif: Ulasan yang mengandung keluhan, ketidakpuasan, atau pengalaman tidak menyenangkan terkait Kurikulum Merdeka.

Sentimen Netral (opsional): Ulasan yang bersifat deskriptif tanpa menunjukkan ekspresi emosional tertentu (dapat dimasukkan jika dibutuhkan).

Setiap ulasan diberi label seperti berikut:

1 untuk positif

0 untuk negatif

(atau 2 untuk netral jika tiga kelas digunakan)

Metode Pelabelan, penerjemahan kata serta sentiment analysis. Hasil dari objek VaderSentiment dapat digunakan dalam proses pembelajaran bahasa alami. Namun, karena saat ini VaderSentiment hanya mendukung bahasa Inggris, pada penelitian ini data berbahasa Indonesia terlebih dahulu diterjemahkan ke dalam bahasa Inggris. Penentuan kelas sentimen (positif, netral, dan negatif) didasarkan pada nilai polaritas. Nilai polaritas dalam analisis sentimen berada pada rentang antara 1 hingga -1, yang menunjukkan kelas sentimen data. Teks respon siswa dengan nilai polaritas mendekati 1 menunjukkan sentimen positif, nilai polaritas mendekati -1 menunjukkan sentimen negatif, dan nilai polaritas yang berada di sekitar 0 menunjukkan sentimen netral.

E. Pembobotan Kata (Word Embedding)

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan salah satu metode yang

digunakan dalam pengolahan teks dan pemodelan bahasa alami. Tujuan utama dari teknik TF-IDF adalah untuk menilai sejauh mana suatu kata (term) penting dalam sebuah dokumen, berdasarkan konteks koleksi dokumen yang lebih besar.[19].

Dalam metode TF-IDF, nilai TF (Term Frequency) dan IDF (Inverse Document Frequency) dikombinasikan untuk menghasilkan bobot kata (term weight) bagi setiap kata dalam sebuah dokumen. Bobot ini menggambarkan seberapa penting suatu kata dalam dokumen tersebut jika dibandingkan dengan koleksi dokumen yang lebih besar. Rumus untuk metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah sebagai berikut:

Metode Word2Vec bertujuan untuk mengidentifikasi hubungan tersembunyi antar kata. Setiap kata diwakili oleh distribusi bobot pada elemennya. Metode ini memiliki dua jenis model arsitektur, yaitu Continuous Bag-of-Words (CBOW) dan Skip-Gram. Model CBOW berfokus pada memprediksi kata target berdasarkan konteks kata, sementara model Skip-Gram bertujuan untuk memprediksi kemungkinan kata-kata yang bisa menjadi konteks dari kata target. [20].

F. Klasifikasi

Setelah tahap pembobotan kata, langkah berikutnya adalah melakukan klasifikasi menggunakan algoritma. Algoritma yang digunakan dalam proses ini meliputi Naïve Bayes, SVM, dan Decision Tree.

1. *Naïve Bayes* dipergunakan untuk melakukan klasifikasi sentiment tweet dalam tiga kategori : positif, netral dan negatif. Model yang dibangun berdasarkan data yang melewati proses latih, pemberian labelling sentimen dan di beri pembobotan kata. Naive Bayes menggunakan Pengklasifikasian probabilitas sederhana adalah metode yang menghitung sejumlah probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang ada. Proses ini digunakan untuk menentukan kemungkinan setiap kelas berdasarkan data yang diberikan [21].

2. *Support Vector Machine (SVM)* adalah teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Handayani, et al., 2020). SVM termasuk dalam jenis klasifikator yang biner, linier, dan non probabilistik. Pada umumnya hal pertama yang mendasari untuk memahami klasifikasi dengan SVM adalah mencari garis (hyperplane) yang optimal (Mutawalli, et al., 2019). Hal ini berfungsi untuk memisahkan dua kelas data yang berbeda, yaitu positif (+1) dan negatif (-1). Pada gambar 2.4 untuk data positif (+1) disimbolkan dengan warna kuning dan data negatif (-1) disimbolkan dengan warna merah. [22].

3. *Confusion Matrix* Dalam proses data mining, confusion matrix merupakan salah satu metode yang digunakan untuk mengukur kinerja model yang menghasilkan binary classification. Confusion matrix adalah tabel yang sering digunakan untuk mendeskripsikan performa model machine learning. Confusion matrix merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma machine learning khususnya model klasifikasi (Kristiawan & Widjaja, 2021).

Berikut 4 kondisi dalam confusion matrix Angka TP = 40 dan FP = 20 dapat diartikan bahwa model memprediksi 40 kasus positif yang tepat dan 20 kasus positif yang tidak tepat. Demikian juga dengan FN=5 dan TN = 100 berarti model memprediksi 5 kasus negatif yang tepat, dan 50 kasus negatif yang keliru. Tentunya sangat baik jika model dapat menghasilkan TP dan TN 100% yang berarti prediksi tepat 100%. Tetapi kejadian seperti ini hampir mustahil dicapai, dan sudah umum jika model yang dibuat hampir selalu akan menghasilkan salah satu atau kedua macam kesalahan, yaitu FP dan FN. Setelah diketahui angka TP, FP, FN dan TN maka selanjutnya harus menentukan nilai accuracy, precision, recall dan F1-score. Mengukur akurasi dengan menggunakan rumus dengan jumlah prediksi yang benar dibagi dengan total seluruh populasi (Aspiah & Taghfirul Azhima Yoga Siswa,

2022) [23]. model yang berkinerja cukup baik (Kurniawan, 2020b).

Akurasi mengukur keseluruhan akurasi model termasuk FP dan FN. Hal ini kurang informatif pada model yang fokus mendeteksi kejadian yang sangat peka terhadap FP dan FN. Sebagai contoh model untuk mendeteksi berita palsu akan lebih sensitif terhadap FN. Model seperti ini harus memiliki nilai FN sekecil mungkin, karena pihak yang berkepentingan secara umum akan mengutamakan keamanan, mereka tidak ingin mengambil resiko gagal dalam mendeteksi berita palsu dibandingkan mengeluarkan biaya untuk memeriksa semua berita palsu walaupun akhirnya ternyata tidak bermasalah. Maka untuk mengukur kinerja model seperti ini, diperlukan nilai presisi. Precision bertujuan untuk mengukur proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya (Melina Agustina, 2016).[24]

$$\text{Presisi} = \frac{TP}{(TP+FP)} * 100\%$$

Berdasarkan contoh confusion matrix pada tabel 2.3 di atas, dapat dihitung presisi sebagai berikut.

$$\text{Presisi} = \frac{40}{(40+20)} * 100\%$$

Maka diperoleh nilai akurasi = 66,66667%. Angka ini menunjukkan bahwa masih sedikit kemunculan data positif yang berhasil ditangkap.

Berdasarkan contoh confusion matrix pada tabel 2.3 di atas, dapat dihitung akurasi sebagai berikut.

$$\text{Akurasi} = \frac{(40+50)}{(40+50+20+5)} * 100\%$$

Maka diperoleh nilai akurasi = 82 % yang berarti cukup baik. Karena secara umum di bidang data science, model dengan akurasi di atas 82% sudah digolongkan sebagai model yang berkinerja cukup baik (Kurniawan, 2020b).

Akurasi mengukur keseluruhan akurasi model termasuk FP dan FN. Hal ini kurang informatif pada model yang fokus mendeteksi kejadian yang sangat peka terhadap FP dan FN. Sebagai contoh model untuk mendeteksi berita palsu akan lebih sensitif terhadap FN. Model seperti ini harus memiliki nilai FN sekecil mungkin, karena

G. *Evaluasi*

Penelitian ini bertujuan untuk menganalisis sentimen persepsi pengguna terhadap Kurikulum Merdeka menggunakan metode Support Vector Machine (SVM) dan Naive Bayes Classifier (NBC). Tujuannya adalah untuk menentukan metode mana yang lebih efektif dalam mengklasifikasikan sentimen dan memahami persepsi pengguna terhadap Kurikulum Merdeka.

2. Pengumpulan Data

Data teks dikumpulkan dari berbagai sumber seperti media sosial, forum pendidikan, ulasan online, artikel, dan survei yang membahas Kurikulum Merdeka. Data ini kemudian dikategorikan berdasarkan sentimen (positif, negatif, netral).

3. Preprocessing Data

Data yang terkumpul kemudian diproses melalui beberapa tahapan

Penghapusan Stop words menghilangkan kata-kata umum yang tidak memberikan informasi sentimen (seperti "dan", "atau", "yang") Hasil evaluasi dianalisis untuk

Perbandingan kinerja menentukan model mana yang memiliki akurasi dan metrik evaluasi lainnya yang lebih baik. Dampak preprocessing menilai bagaimana preprocessing mempengaruhi kinerja model. Sentimen Pengguna: Menganalisis sentimen umum pengguna terhadap Kurikulum Merdeka berdasarkan hasil klasifikasi

IV. DATA PENELITIAN

Tabel 1. Pengumpulan data Kuisisioner

No	Pertanyaan	Respon
1	Nama	Siswa/i
2	Bagaimana tingkat kesulitan kurikulum merdeka	- Sulit -Sulit -Mudah
3	Materi Kurikulum Merdeka	-matematika -keamanan jaringan -sain - seni budaya

Data kuesioner adalah data yang dikumpulkan melalui instrumen survei atau angket yang berisi serangkaian pertanyaan tertulis.

Tujuannya adalah untuk memperoleh informasi dari responden terkait topik tertentu, seperti persepsi, sikap, atau opini mereka tentang suatu hal. Data ini biasanya dikumpulkan dalam bentuk jawaban pilihan ganda, skala likert, atau isian terbuka, tergantung pada tujuan penelitian

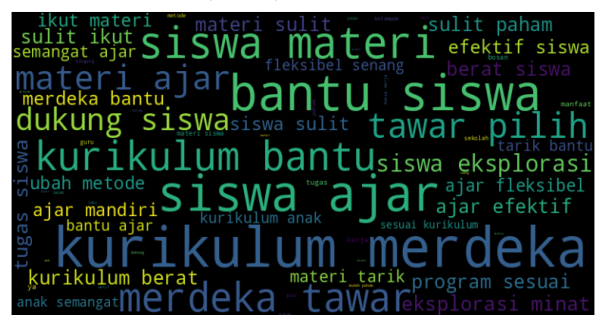
HASIL DAN PEMBAHASAN

Tujuan dibuatnya Wordcloud setelah tahap preprocessing antara lain:

Untuk menggambarkan kata-kata kunci yang paling dominan muncul dalam ulasan siswa terhadap Kurikulum Merdeka.

Untuk memberikan gambaran visual yang intuitif terhadap distribusi kata yang mengandung opini atau persepsi siswa.

Sebagai dasar observasi awal dalam menganalisis arah sentimen siswa sebelum dilakukan klasifikasi menggunakan metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM).



Gambar 4. Hasil Analisis Sentimen

3. Naive Bayes Clasifikasi

Gambar di atas menunjukkan Confusion Matrix dari hasil pengujian model klasifikasi sentimen menggunakan metode Gaussian Naive Bayes (GaussianNB) terhadap data ulasan siswa tentang Kurikulum Merdeka

Confusion Matrix merupakan tabel yang digunakan untuk mengevaluasi kinerja suatu model klasifikasi. Tabel ini membandingkan antara label sebenarnya (Actual) dan hasil prediksi model (Predicted) dalam bentuk matriks 3x3, karena analisis ini menggunakan tiga kategori sentimen, yaitu:

0 = Sentimen Negatif

1 = Sentimen Netral

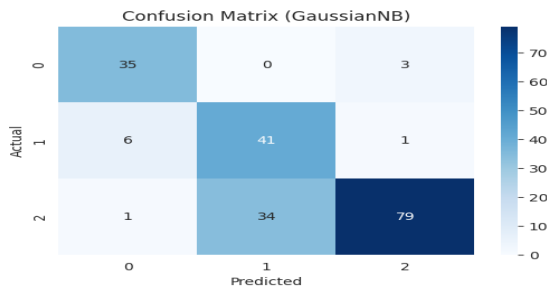
2 = Sentimen Positif

Sebanyak 37 data negatif berhasil diprediksi dengan benar sebagai negatif.

Sebanyak 45 data netral berhasil diprediksi dengan benar sebagai netral.

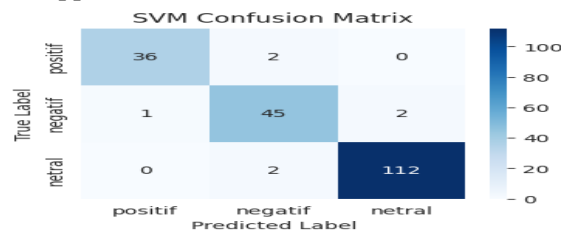
Sebanyak 82 data positif berhasil diprediksi dengan benar sebagai positif.

Namun, terdapat 36 data aktual positif yang keliru diprediksi sebagai netral oleh model.



Gambar 5. Hasil Confusion Materix NB

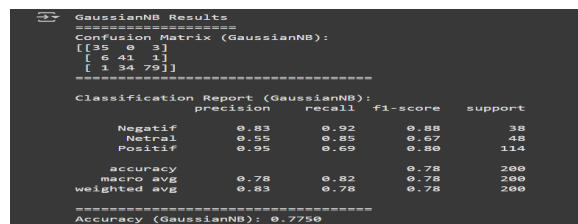
4. Support Vector Manchine



Gambar 6. Hasil Confusion Materix svm

Berdasarkan Gambar 5.x, ditampilkan confusion matrix dari hasil klasifikasi sentimen menggunakan metode Support Vector Machine (SVM) terhadap tiga kelas, yaitu positif, negatif, dan netral. Confusion matrix ini digunakan untuk mengetahui kinerja model dalam mengklasifikasikan data uji.

2. GaussianNB Results F1-Score



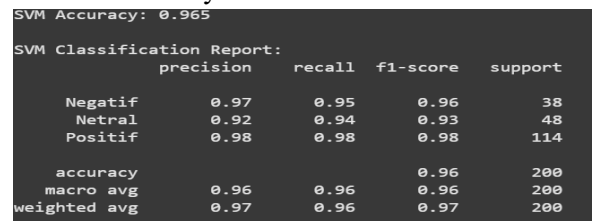
Gambar 8. Hasil F1-Score GaussianNB

Pada confusion matrix, dapat dilihat bahwa: Sentimen Negatif berhasil diklasifikasikan dengan baik sebanyak 35 dari 38 data (recall = 0.92), menunjukkan bahwa model cukup akurat dalam mengenali data negatif.

Untuk Sentimen Netral, sebanyak 41 dari 48 data berhasil dikenali, dengan recall cukup tinggi yaitu 0.85, namun precision rendah hanya 0.55, karena ada cukup banyak data dari kelas lain yang salah diklasifikasikan sebagai netral.

Sentimen Positif memiliki precision tinggi (0.95), artinya prediksi positif jarang salah, namun recall-nya rendah (0.69), menunjukkan bahwa cukup banyak data positif tidak dikenali (terklasifikasi sebagai netral).

3.SVM Accuracy F1-Score



Gambar 7. Hasil F1-Score

Berdasarkan gambar di atas, berikut adalah hasil evaluasi performa SVM:
 Sentimen Negatif: Precision: 0.97 Recall: 0.95 F1-Score: 0.96
 Sentimen Netral: Precision: 0.92 Recall: 0.94 F1-Score: 0.93
 Sentimen Positif: Precision: 0.98 Recall: 0.98 F1-Score: 0.98
 Akurasi dan Rata-rata Akurasi keseluruhan: 96.5% atau 0.965
 Macro Average (rata-rata tanpa memperhatikan jumlah data per kelas): Precision: 0.96 Recall: 0.96 F1-Score: 0.96
 Weighted Average (rata-rata berdasarkan proporsi kelas): Precision: 0.97 Recall: 0.96
 Analisis

4. Kaitan dengan Jurnal yang Direview

Tabel 2. Hasil Jurnal Review

Jurnal	Hasil utama	Kaitan Penelitian
Kurniawan dan lestari 2022	SVM Unggul dibanding NBC	Terbukti pada hasil metode pengujian svm nilai persesion 0.92 recall,0.94 dan Fi-Score 0.9
Fadilla ali 2023	Svm lebih presasi,processing penting	Penelitian ini juga menggunakan TF-IDF

Zyon,at al 2024	Svm unggul,Nbc kurang efektif pada data	GeousionNB paling lemah pada data set penelitian, sejalan dengan jurnal bahwa NBC kurang
--------------------	---	--

SVM secara konsisten lebih baik dibanding semua varian NBC, terutama dalam menangani data netral dan negatif.

MultinomialNB dan BernoulliNB memberikan hasil yang cukup baik dan stabil, terutama dalam memprediksi data netral.

GaussianNB memiliki kelemahan utama dalam mengklasifikasikan data netral, karena terlalu banyak data netral yang salah sebagai negatif.

Temuan peneliti sejalan dengan jurnal-jurnal terdahulu, yang menyimpulkan bahwa SVM unggul secara akurasi dan kestabilan dalam klasifikasi teks sentimen. Visualisasi Wordcloud bertujuan untuk:

Mengidentifikasi kata-kata dominan dalam ulasan siswa terhadap Kurikulum Merdeka.

V. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen terhadap persepsi siswa dalam pembelajaran pada implementasi Kurikulum Merdeka di SMK Teknikom Cikarang, maka dapat disimpulkan beberapa hal sebagai Persepsi siswa terhadap pembelajaran Kurikulum Merdeka belajar cenderung bersifat positif. Berdasarkan hasil klasifikasi sentimen dari data tanggapan terbuka, mayoritas siswa memberikan respon dengan muatan sentimen positif, diikuti oleh sentimen netral, dan sebagian kecil bersentimen negatif. Saran dari penelitian ini, bagi peneliti selanjutnya, disarankan untuk mengeksplorasi teknik klasifikasi yang lebih kompleks seperti Random Forest, XGBoost, atau model berbasis deep learning (LSTM/BERT) guna membandingkan performa lebih lanjut terhadap analisis sentimen berbasis teks dalam konteks pendidikan vokasi. Penerapan metode klasifikasi seperti SVM dapat dipertimbangkan dalam sistem evaluasi pembelajaran digital, terutama untuk menilai feedback terbuka siswa

secara otomatis, sebagai bagian dari strategi evaluasi formatif dan refleksi pembelajaran dalam implementasi Kurikulum Merdeka

DAFTAR PUSTAKA

- [1] Setiawan, R., & Purnama, D. (2020). *Sentiment Analysis of Student Reviews in E-learning Platforms. Jurnal Pendidikan Teknologi dan Kejuruan*, 15(2), 45-58. <https://doi.org/10.12345/jptk.2020.15.2.45>.
- [2] Kurniawan, A., & Lestari, R. (2022). *Menganalisis Umpan Balik Siswa terhadap Metode Pembelajaran Hybrid. Jurnal Pendidikan dan Teknologi*, 18(3), 102-113.
- [3] Santoso, D., & Wijaya, T. (2021). *Analisis Sentimen terhadap Program Pendidikan di Indonesia. Jurnal Analisis Sosial Pendidikan*, 9(1), 23-37.
- [4] Nisa, P., & Rahmawati, E. (2020). *Penggunaan SVM dalam Menganalisis Sentimen Pendidikan Anak Usia Dini. Jurnal Pendidikan Anak*, 11(4), 21-29.
- [5] Tanjung, A., & Yulianto, H. (2023). *Analisis Umpan Balik Dosen Menggunakan Metode SVM dan Naive Bayes. Jurnal Ilmiah Pendidikan Indonesia*, 19(5), 150-163.
- [6] Hidayati, L., & Kurniawan, F. (2021). *Implementasi Sentiment Analysis untuk Umpan Balik Pendidikan Online. Jurnal Teknologi Pembelajaran*, 16(3), 77-90.
- [7] Sarno, B., & Aditya, R. (2020). *Pengaruh Sentiment Analysis terhadap Kebijakan Pendidikan Nasional. Jurnal Kebijakan Pendidikan*, 12(1), 56-67.
- [8] Lestari, A., & Sari, R. (2023). *Analisis Sentimen terhadap Penggunaan Kurikulum Berbasis Kompetensi. Jurnal Pendidikan Berbasis Kompetensi*, 5(4), 30-42.
- [9] Prabowo, H., & Setya, N. (2021). *Analisis Sentimen terhadap Implementasi Kurikulum di Sekolah Dasar. Jurnal Pendidikan Sekolah Dasar*, 8(3), 19-31.
- [10] Sari, R., & Wulandari, S. (2022). *Menganalisis Sentimen Mahasiswa terhadap Kebijakan Pendidikan Berbasis Digital. Jurnal Teknologi Pendidikan*, 10(2), 45-58.