

# Komparasi Performa Klasifikasi Sentimen Masyarakat Terhadap Kurikulum Merdeka di Sekolah Menggunakan SVM dan KNN

Risa Fitria Apriyani, Dyah Ayu Megawaty\*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: <sup>1</sup>risa\_fitria\_apriyani@teknokrat.ac.id, <sup>2,\*</sup>dyahayumegawaty@teknokrat.ac.id

Email Penulis Korespondensi: dyahayumegawaty@teknokrat.ac.id

Submitted: 30/01/2025; Accepted: 28/03/2025; Published: 28/03/2025

**Abstrak**—Kurikulum Merdeka merupakan kebijakan pendidikan strategis yang bertujuan meningkatkan fleksibilitas pembelajaran serta pengembangan kompetensi peserta didik di era abad ke-21. Penelitian ini berfokus pada analisis sentimen masyarakat terhadap implementasi Kurikulum Merdeka menggunakan dua algoritma pembelajaran mesin, yaitu Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Salah satu tantangan utama dalam penelitian ini adalah ketidakseimbangan data sentimen yang mencakup kelas negatif, netral, dan positif. Untuk mengatasi hal ini, diterapkan teknik Synthetic Minority Oversampling Technique (SMOTE) guna menyeimbangkan distribusi data antar kelas. Hasil penelitian menunjukkan bahwa metode SVM unggul dibandingkan KNN dengan akurasi keseluruhan mencapai 92% dan F1-score tinggi pada kelas mayoritas (Netral: 96%), meskipun performa pada kelas minoritas (Negatif: 43% dan Positif: 40%) masih memerlukan peningkatan. Sebaliknya, metode KNN mencatat akurasi keseluruhan lebih rendah sebesar 31% namun memiliki distribusi kesalahan yang lebih merata. Setelah penerapan SMOTE, terjadi peningkatan signifikan pada kedua metode, terutama dalam mengenali kelas minoritas. Penelitian ini menyimpulkan bahwa SVM lebih efektif untuk tugas klasifikasi sentimen pada dataset dengan ketidakseimbangan kelas, dan merekomendasikan eksplorasi lebih lanjut terhadap metode ensemble guna meningkatkan kualitas prediksi dan generalisasi model.

**Kata Kunci:** Klasifikasi Sentimen; KNN; Kurikulum Merdeka; SMOTE; SVM

**Abstract**—The Independent Curriculum is a strategic education policy that aims to increase learning flexibility and develop student competencies in the 21st century. This research focuses on analyzing public sentiment towards the implementation of the Independent Curriculum using two machine learning algorithms, namely Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). One of the main challenges in this study is the imbalance of sentiment data that includes negative, neutral, and positive classes. To overcome this, the Synthetic Minority Oversampling Technique (SMOTE) technique was applied to balance the distribution of data between classes. The results show that the SVM method is superior to KNN with an overall accuracy of 92% and a high F1-score in the majority class (Neutral: 96%), although the performance in the minority class (Negative: 43% and Positive: 40%) still needs improvement. In contrast, the KNN method recorded a lower overall accuracy of 31% but had a more even distribution of errors. After the implementation of SMOTE, there was a significant improvement in both methods, especially in recognizing minority classes. This study concludes that SVM is more effective for sentiment classification tasks on datasets with class imbalances, and recommends further exploration of ensemble methods to improve the quality of prediction and model generalization.

**Keywords:** Sentiment Classification; KNN; Independent Curriculum; SMOTE; SVM

## 1. PENDAHULUAN

Di Indonesia, sistem pendidikan telah mengalami banyak perubahan seiring perkembangan kebutuhan masyarakat dan tantangan global. Salah satu langkah besar dalam pengelolaan pendidikan di Indonesia adalah pengembangan kurikulum yang disesuaikan dengan dinamika zaman. Kurikulum yang efektif tidak hanya berfungsi sebagai pedoman pembelajaran, tetapi juga sebagai alat untuk mewujudkan pendidikan yang relevan dengan kebutuhan dunia kerja, sosial, dan budaya. Dalam perkembangannya, sistem pendidikan di Indonesia telah mengalami berbagai pembaruan kurikulum untuk menjawab kebutuhan masyarakat yang terus berkembang. Kurikulum tidak hanya sekadar alat pedagogis, tetapi juga strategi kebijakan dalam menciptakan generasi yang cerdas, berkarakter, dan adaptif terhadap dinamika global[1].

Kurikulum Merdeka adalah salah satu kebijakan transformasi pendidikan terbaru yang diperkenalkan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Indonesia. Kurikulum ini dirancang untuk memberikan kebebasan kepada sekolah dalam mengembangkan proses pembelajaran yang relevan dengan karakteristik dan potensi peserta didik. Pendekatan dalam Kurikulum Merdeka lebih menekankan pada pembelajaran berbasis proyek (project-based learning), yang bertujuan untuk mengembangkan kompetensi esensial sekaligus membangun karakter peserta didik. Kurikulum ini juga memberi ruang bagi sekolah dan guru untuk fleksibel dalam menentukan metode dan materi pembelajaran sesuai kebutuhan lokal. Begitu pentingnya kurikulum dalam bidang pendidikan karena menjadi suatu alat, acuan, dasar atau pandangan hidup[2].

Telah banyak penelitian yang dilakukan mengenai analisis sentimen dengan data pengguna Twitter, seperti: Komparasi Metode Klasifikasi Untuk Analisis Sentimen Pengguna Twitter Terhadap Penerapan Kurikulum Merdeka. Metode yang digunakan untuk klasifikasi adalah k-Nearest Neighbor (K-NN), Naive Bayes (NB) dan Support Vector Machine (SVM). Komparasi metode ini bertujuan untuk menemukan metode terbaik dari dataset yang ada, berdasarkan hasil akurasi yang dihasilkan dari ketiga metode tersebut. Dataset tweet yang dikumpulkan sebanyak 1000 tweet, setelah data dibersihkan diperoleh 220 tweet yang terdiri dari 110 tweet positif dan 110 tweet negatif. Berdasarkan perhitungan ketiga metode dengan menggunakan tools Rapidminer didapatkan akurasi sebagai berikut:

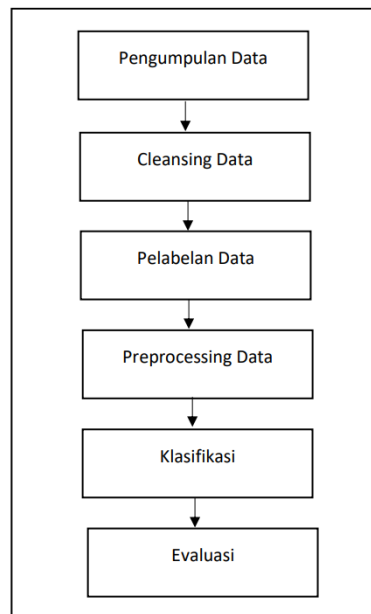
untuk metode K-NN adalah 76,36%, akurasi metode NB adalah 70,91%, sedangkan akurasi metode SVM adalah 62,73%. [3]

Penelitian yang dilakukan oleh Wachid Darmawan dan rekan rekannya yang berjudul Analisis Sentimen Penerapan Kurikulum Merdeka Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor Dengan Forward. Berdasarkan eksperimen yang sudah dilakukan metode K-NN akurasinya sebesar 73.64%, sedangkan K-NN + FS akurasinya meningkat menjadi 76.82%. berdasarkan perhitungan machine learning dihasilkan Sentimen Analisis bernada atau opini negatif yang banyak di tweetkan oleh masyarakat terkait penerapan Kurikulum Merdeka [4].

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap implementasi Kurikulum Merdeka menggunakan pendekatan analisis sentimen berbasis pembelajaran mesin. Dengan memanfaatkan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN), penelitian ini mengklasifikasikan sentimen masyarakat menjadi kategori positif, negatif, dan netral. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam memahami persepsi masyarakat terhadap Kurikulum Merdeka, sehingga dapat digunakan sebagai dasar evaluasi kebijakan. Selain itu, penelitian ini juga menjadi referensi dalam pemilihan metode analisis sentimen yang tepat untuk memahami respons masyarakat terhadap kebijakan pendidikan secara luas. Dengan demikian, penelitian ini tidak hanya relevan secara akademis, tetapi juga memberikan dampak praktis dalam mendukung pengembangan sistem pendidikan yang lebih inklusif dan adaptif di Indonesia.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian



**Gambar 1.** Tahap Penelitian

Penelitian ini dimulai dengan pengumpulan data dari media sosial atau survei publik terkait implementasi Kurikulum Merdeka. Setelah data terkumpul, dilakukan cleansing data untuk menghapus informasi yang tidak relevan atau tidak lengkap. Kemudian, tahap pelabelan data dilakukan dengan mengategorikan sentimen menjadi tiga kelas: positif, negatif, dan netral. Selanjutnya, dilakukan preprocessing data, seperti case folding, tokenisasi, penghapusan stopwords, normalisasi, dan stemming.

Setelah itu, tahap klasifikasi model diterapkan menggunakan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Tahapan terakhir adalah evaluasi model, yang mengukur akurasi, precision, recall, dan F1-score untuk menilai kinerja model dalam mengklasifikasikan sentimen masyarakat terhadap Kurikulum Merdeka, dengan penerapan teknik SMOTE untuk menangani ketidakseimbangan kelas dalam data. Struktur tahap penelitian dapat dilihat pada Gambar 1.

### 2.2 Pengumpulan Data

Data dikumpulkan dengan menggunakan teknik crawling seperti Tweepy untuk mengunduh data dari Twitter. Kata kunci pencarian yang digunakan seperti “Kurikulum Merdeka”, “Implementasi Kurikulum Merdeka”, dan “Kurikulum Merdeka Sekolah”, “Kurikulum Merdeka Belajar Sekolah” dan “Merdeka Belajar Sekolah”. Pada tahap crawling, data mentah yang digunakan diambil dari Google Play Store pada laman aplikasi Flip dengan teknik scraping [5]. Proses ini menghasilkan sebanyak 6.362 data berupa tweet dalam bahasa Indonesia. Data yang terkumpul terdiri dari opini masyarakat yang dipublikasikan antara tahun 2023 hingga 2024.

### 2.3 Data Cleansing

Data cleansing adalah proses penghilangan komponen tertentu yang terdapat dalam tweet yakni Uniform Resource Locator (URL), username, RT (Retweet), karakter HTML, dan hashtag[6]. Proses ini penting dilakukan untuk memastikan bahwa data yang dianalisis benar-benar merepresentasikan opini atau isi utama dari pengguna, tanpa gangguan dari elemen-elemen teknis atau simbolik yang tidak relevan.

### 2.4 Pelabelan Data

Pelabelan dengan VADER Lexicon menggunakan kamus kata-kata yang telah diberi skor sentimen untuk menentukan tingkat sentimen teks (positif, negatif, atau netral). VADER menghitung sentimen berdasarkan kata-kata polaritas, intensitas (misalnya, penguat seperti "very"), tanda baca, huruf kapital, dan memberikan skor komposit (compound) dalam rentang -1 (sangat negatif) hingga +1 (sangat positif). Teks dengan compound > 0.05 dilabeli positif, < -0.05 sebagai negatif, dan di antara keduanya netral.

### 2.5 Data Preprocessing

Data Preprocessing adalah proses untuk mempersiapkan data mentah agar siap digunakan dalam analisis atau model pembelajaran mesin. Tahap ini melibatkan proses pengolahan awal data teks, termasuk tokenisasi (mengubah teks menjadi token atau kata-kata), penghapusan stopwords (kata-kata umum yang tidak memberikan makna penting), dan stemming (mengubah kata-kata menjadi kata dasar) [7].

### 2.6 Proses Klasifikasi

Klasifikasi merupakan suatu proses pengkategorian yang dilakukan terhadap sekumpulan data. Proses tahap di mana model yang telah dilatih digunakan untuk mrngkategorikan data baru ke dala kelas- kelas tertentu, seperti sentimen positif, negatif, atau netral. Sebelum proses kalsifikasi, data dibagi menjadi 2 bagian, yakni: data training dan data testing, dengan komposisi pembagia datan testing 20 dari total dataset yang ada[8].

### 2.7 Support Vector Machine

SVM merupakan salah satu metode dalam pattern recognition yang bekerja atas prinsip menemukan hyperplane terbaik yang memisahkan dua buah class pada input space dan mencari titik maksimalnya[9]. SVM berusaha menemukan garis atau bidang yang memisahkan data dengan jarak terbesar antara kelas-kelas tersebut. SVM diperkenalkan oleh Vapnik sebagai model machine learningberbasis kernel untuk klasifikasi dan regresi[10]

$$y_i(w \cdot x_i + b) \geq 1 \quad (1)$$

$$y_i(w \cdot x_i + b) \leq -1 \quad (2)$$

Bobot  $w$  merupakan parameter yang menunjukkan arah dari *hyperplane* pemisah dalam ruang fitur, sedangkan  $x_i$  adalah vector fitur dari kata ke  $-i$ . Nilai  $b$  berfungsi sebagai bias yang menggeser posisi hyperplane agar tidak selalu melalui titik asal. Setiap data  $x_i$  memiliki label  $y_i$  kelas dimana nilai  $y_i = 1$  menunjukkan bahwa data tersebut dakam kelas positif, dan  $y_i = -1$  menunjukkan bahwa data termasuk dalam kelas negatif.

### 2.8 K-Nearest Neighbor (KNN)

Metode ini merupakan salah satu metode yang digunakan untuk melakukan klasifikasi. KNN adalah suatu algoritma pembelajaran mesin bekerja dengan cara menghitung jarak antara data uji dengan data latih[11]. Penggunaan K-Nearest Neighbor bertujuan untuk mengklasifikasikan objek baru berdasarkan data pembelajaran yang paling dekat dengan objek baru tersebut[12]. KNN mudah diimplementasikan dan tidak memerlukan pelatihan model yang rumit, tetapi dapat menjadi sangat lambat jika datasetnya besar karena memerlukan perhitungan jarak terhadap setiap titik data dalam dataset.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Jarak *Euclidean* antara dua titik  $x$  dan  $y$  yang dilambangkan sebagai merupakan ukuran jarak berdasarkan selisih kuadrat antara nilai-nilai fitur pada masing-masing dimensi. Pada setiap dimensi ke  $-i$ ,  $x_i$  dan  $y_i$  mewakili koordinat atau nilai fitur dari titik  $x$  dan  $y$ . Selisish kuadrat antara  $x_i$  dan  $y_i$  dihitung dengan rumus  $(x_i - y_i)$ . Jarak Euclidean dihitung dengan menjumlahkan seluruh seilish kuadrat ini untuk dimensi yang dimiliki oleh data.

### 2.9 SMOTE

Metode SMOTE (Synthetic Minority Oversampling Technique) adalah salah satu metode yang begitu populer untuk digunakan dalam mengatasi sebuah ketidakseimbangan kelas[13]. Teknik ini membantu meningkatkan distribusi data kelas minoritas tanpa hanya menduplikasi data asli, sehingga mengurangi bias model terhadap kelas mayoritas. Data sintetis tersebut dihasilkan dengan cara mengambil titik data dari kelas minoritas dan membuat sampel baru di antara titik tersebut dan tetangga terdekatnya



### 2.10 Evaluasi

Tahap evaluasi adalah proses untuk menilai kinerja model machine learning dalam menyelesaikan tugas tertentu, seperti klasifikasi atau prediksi. Pada tahap ini, model yang telah dilatih diuji menggunakan data uji untuk mengukur seberapa baik model dapat menggeneralisasi pola yang telah dipelajari. Evaluasi ini dilakukan dengan membagi data menjadi set pelatihan dan pengujian, kemudian mengukur performa model menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Pemilihan metrik evaluasi yang digunakan bergantung pada kompleksitas dan distribusi data yang dimiliki [14].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1 = 2 \times \frac{Precision+Recall}{Precision+Recall} \tag{6}$$

True Positive (TP) adalah jumlah data yang benar-benar diklasifikasikan sebagai positif oleh model, yang artinya prediksi model sesuai dengan kondisi sebenarnya. False Positive (FP) mengacu pada data yang sebenarnya negatif, tetapi salah diprediksi sebagai positif oleh model. Di sisi lain, True Negative (TN) adalah data yang benar-benar negatif dan berhasil dikenali dengan benar oleh model, sementara False Negative (FN) adalah data positif yang justru salah diklasifikasikan sebagai negatif oleh model. Kesalahan-kesalahan ini penting untuk dievaluasi dalam menentukan akurasi dan efektivitas model klasifikasi.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan Teknik web scraping untuk mengambil opini masyarakat terkait Kurikulum Merdeka. Data dikumpulkan dari sumber media social X, proses ini menghasilkan data sebanyak tweet yang relevan dengan topik yang dicari. Data ini kemudian disimpan dalam format CSV untuk pemrosesan lebih lanjut. Hasil dari proses ini dapat dilihat pada Tabel 1 berikut ini:

**Tabel 1.** Hasil Pengumpulan Data

Tweet
@Lone_Lynx_ Tutup aj sekolah lah...KALO pemerintah gak mau kasih kurikulum yg jelas..itu tidur termasuk kurikulum merdeka belajar ya? Ya gak salah kalo masuk. Tidak ada perlindungan guru yg jelas. Masih ada guru honorer(harusnya pengajar standar gaji sama kecuali yg punya jabatan)

Tabel 1 menyajikan hasil pengumpulan data berupa cuitan dari media sosial X yang mencerminkan opini masyarakat terkait Kurikulum Merdeka. Salah satu cuitan memperlihatkan adanya kritik terhadap implementasi kurikulum yang dianggap belum jelas arah dan perlindungannya, khususnya terhadap guru honorer.

### 3.2 Data Cleansing

Data yang diperoleh dari web scraping kemudian disproses melalui tahap pembersihan. Tahap cleansing untuk menghilangkan karakter atau simbol, emoticon, dan link url yang tidak memiliki hubungan informasi[15]. Hasil dari proses Cleansing dapat dilihat pada Tabel 1 berikut ini:

**Tabel 2.** Data Hasil *Cleansing*

Proses	Hasil
Sebelum	@Lone_Lynx_ Tutup aj sekolah lah...KALO pemerintah gak mau kasih kurikulum yg jelas..itu tidur termasuk kurikulum merdeka belajar ya? Ya gak salah kalo masuk. Tidak ada perlindungan guru yg jelas. Masih ada guru honorer(harusnya pengajar standar gaji sama kecuali yg punya jabatan).
Sesudah	Lone Lynx Tutup aj sekolah lah KALO pemerintah gak mau kasih kurikulum yg jelasitu tidur termasuk kurikulum merdeka belajar ya Ya gak salah kalo masuk Tidak ada perlindungan guru yg jelas Masih ada guru honorerharusnya pengajar standar gaji sama kecuali yg punya jabatan

Tabel 2 menunjukkan hasil proses data cleansing pada data yang diperoleh melalui web scraping. Sebelum proses cleansing, data mengandung simbol, emoticon, dan link yang tidak relevan. Proses cleansing bertujuan menghilangkan elemen-elemen ini untuk memastikan hanya data yang memiliki informasi penting yang tersisa. Hal ini diperlukan agar data menjadi lebih bersih dan siap digunakan dalam analisis lebih lanjut.

Setelah proses cleansing, data menjadi lebih terstruktur dan mudah dipahami. Karakter atau simbol yang tidak relevan telah dihapus, sehingga informasi yang tersisa lebih jelas. Dengan data yang sudah bersih, analisis yang

dilakukan menjadi lebih akurat dan dapat memberikan hasil yang lebih bermanfaat. Proses ini penting untuk meningkatkan kualitas data yang digunakan dalam penelitian.

### 3.3 Pelabelan Data

VADER Lexicon adalah metode berbasis kamus yang dirancang untuk menganalisis sentimen teks informal, seperti ulasan dan media sosial. VADER akan menghasilkan skor yang dinyatakan dengan compound yaitu gabungan dari masing-masing skor positif, negatif, dan netral[16].

**Tabel 3.** Hasil Pelabelan Data

Sentimen	Hasil
Negatif	soalnya udah kurikulum merdeka hahaha smangat jugak menajalani hari dewasamu ituh
Netral	Nah ini kok gak viral sih? Ini efek Kurikulum Merdeka. Guru dan anak kreatif. Keren! Lanjutkan menginspirasi Bu
Positif	hasil sekolah di rumah selama pandemi pandemi beres ketemunya kurikulum merdeka

Tabel 3 menunjukkan hasil pelabelan data berdasarkan sentimen, dengan kategori negatif, netral, dan positif. Proses ini dilakukan untuk mengklasifikasikan teks berdasarkan reaksi terhadap kurikulum merdeka. Data yang tidak jelas sebelumnya diproses agar dapat dikategorikan dengan tepat. Hal ini memudahkan analisis selanjutnya dan membantu dalam memahami pola opini publik terhadap kebijakan tersebut.

Setelah pelabelan, data menjadi lebih terstruktur dan mudah dianalisis. Sentimen negatif mencerminkan ketidakpuasan terhadap kurikulum merdeka, sementara sentimen positif menunjukkan dukungan. Kategori netral mencakup pendapat yang tidak condong ke arah tertentu tetapi tetap relevan dalam diskusi. Pelabelan ini memberikan wawasan yang lebih jelas mengenai persepsi masyarakat terhadap perubahan yang diterapkan, serta membantu dalam pengambilan keputusan berbasis data.

### 3.4 Preprocessing Data

Serangkaian langkah yang dilakukan untuk mempersiapkan data mentah sebelum digunakan dalam proses analisis atau model pembelajaran mesin. Tujuannya adalah untuk meningkatkan kualitas data sehingga model dapat bekerja lebih efisien dan menghasilkan hasil yang lebih akurat. Proses data preprocessing dilakukan dengan langkah-langkah berikut untuk memastikan data yang digunakan dalam model pembelajaran mesin (SVM dan KNN) berkualitas tinggi.

#### 3.4.1 Case Folding

*Case folding* merupakan proses yang dilakukan untuk menyeragamkan karakter pada data di dalam text preprocessing[17]. Pada tahap ini, tweet diubah menjadi huruf kecil semua dengan menggunakan bahasa pemrograman Python[18]. Tahap ini penting untuk meningkatkan konsistensi data sebelum analisis lebih lanjut, seperti klasifikasi atau text mining.

**Tabel 4.** Data Hasil *Case Folding*

Proses	Hasil
Sebelum	@Lone_Lynx_ Tutup aj sekolah lah...KALO pemerintah gak mau kasih kurikulum yg jelas..itu tidur termasuk kurikulum merdeka belajar ya? Ya gak salah kalo masuk. Tidak ada perlindungan guru yg jelas. Masih ada guru honorer(harusnya pengajar standar gaji sama kecuali yg punya jabatan).
Sesudah	lone lynx tutup aj sekolah lah kalo pemerintah gak mau kasih kurikulum yg jelasitu tidur termasuk kurikulum merdeka belajar ya ya gak salah kalo masuk tidak ada perlindungan guru yg jelas masih ada guru honorerharusnya pengajar standar gaji sama kecuali yg punya jabatan

Tabel 4 menunjukkan hasil dari proses case folding yang bertujuan untuk menyamakan format teks. Sebelum case folding, data mengandung huruf kapital yang tidak konsisten, yang dapat mengganggu analisis. Proses ini mengubah semua huruf menjadi huruf kecil untuk menciptakan keseragaman. Dengan demikian, perbedaan antara kata-kata yang serupa dapat dihindari.

Setelah case folding, data menjadi lebih konsisten dan siap untuk dianalisis lebih lanjut. Semua huruf kapital diubah menjadi huruf kecil, sehingga teks menjadi lebih mudah diproses. Proses ini mengurangi kemungkinan kesalahan yang disebabkan oleh perbedaan penulisan. Dengan hasil case folding yang sudah dilakukan, analisis dapat berjalan lebih efektif dan efisien.

#### 3.4.2 Tokenizing

Tokenizing merupakan proses pemecahan sekumpulan karakter dalam suatu teks kedalam satuan kata, pada proses ini juga dapat menghilangkan karakter pembatas, menghapus angka, dan menghapus tanda baca atau karakter yang bukan merupakan huruf[19]. Hasil dari proses *tokenization* ditunjukkan pada Tabel 4 berikut ini.

**Tabel 5.** Data Hasil *Tokenizing*

Proses	Hasil
Sebelum	@Lone_Lynx_ Tutup aj sekolah lah...KALO pemerintah gak mau kasih kurikulum yg jelas..itu tidur termasuk kurikulum merdeka belajar ya? Ya gak salah kalo masuk. Tidak ada perlindungan guru yg jelas. Masih ada guru honorer(harusnya pengajar standar gaji sama kecuali yg punya jabatan).
Sesudah	['lone', 'lynx', 'tutup', 'aj', 'sekolah', 'lah', 'kalo', 'pemerintah', 'gak', 'mau', 'kasih', 'kurikulum', 'yg', 'jelasitu', 'tidur', 'termasuk', 'kurikulum', 'merdeka', 'belajar', 'ya', 'ya', 'gak', 'salah', 'kalo', 'masuk', 'tidak', 'ada', 'perlindungan', 'guru', 'yg', 'jelas', 'masih', 'ada', 'guru', 'honorerharusnya', 'pengajar', 'standar', 'gaji', 'sama', 'kecuali', 'yg', 'punya', 'jabatan']

Tabel 5 menunjukkan hasil proses tokenizing, yaitu pemisahan teks menjadi unit kata atau token. Sebelum proses ini, teks masih berupa kalimat utuh yang sulit dianalisis karena mengandung berbagai elemen seperti tanda baca dan spasi tidak teratur. Dengan tokenizing, setiap kata dipisahkan secara sistematis sehingga lebih mudah untuk diproses dalam analisis lebih lanjut.

Setelah tokenizing, teks menjadi lebih terorganisir dan siap untuk tahapan analisis berikutnya. Pemisahan kata memungkinkan sistem memahami konteks setiap kata secara lebih akurat, terutama dalam klasifikasi teks atau analisis sentimen. Dengan hasil tokenizing yang lebih bersih, langkah-langkah selanjutnya seperti stemming atau stopword removal dapat dilakukan dengan lebih efektif.

### 3.4.3 Stopword Removal

*Stopword Removal* merupakan proses penghapusan kata yang tidak diperlukan atau tidak memiliki makna seperti kata sambung, kata ganti dan lain-lain[20]. Tujuan dari proses ini mengurangi kebisingan dalam data dan memungkinkan model focus pada kata-kata relevan untuk meningkatkan efisiensi dan akurasi analisis. Berikut hasil dari proses Stopword Removal dapat dilihat pada Tabel 6.

**Tabel 6.** Data Hasil *Stopword Removal*

Proses	Hasil
Sebelum	@Lone_Lynx_ Tutup aj sekolah lah...KALO pemerintah gak mau kasih kurikulum yg jelas..itu tidur termasuk kurikulum merdeka belajar ya? Ya gak salah kalo masuk. Tidak ada perlindungan guru yg jelas. Masih ada guru honorer(harusnya pengajar standar gaji sama kecuali yg punya jabatan).
Sesudah	['lone', 'lynx', 'tutup', 'aj', 'sekolah', 'kalo', 'pemerintah', 'gak', 'kasih', 'kurikulum', 'yg', 'jelasitu', 'tidur', 'kurikulum', 'merdeka', 'belajar', 'ya', 'ya', 'gak', 'salah', 'kalo', 'masuk', 'perlindungan', 'guru', 'yg', 'guru', 'honorerharusnya', 'pengajar', 'standar', 'gaji', 'kecuali', 'yg', 'jabatan']

Tabel 6 menunjukkan hasil dari proses stopword removal, yaitu penghapusan kata-kata yang tidak memiliki makna signifikan dalam analisis teks. Sebelum proses ini, teks masih mengandung kata-kata seperti kata sambung dan kata ganti yang tidak relevan untuk analisis. Stopword removal bertujuan untuk mengurangi kebisingan dalam data agar lebih fokus pada informasi penting. Dengan demikian, hasil analisis menjadi lebih akurat dan efisien.

Setelah stopword removal, teks menjadi lebih bersih dan hanya menyisakan kata-kata yang memiliki makna signifikan. Kata-kata yang dihapus umumnya tidak berkontribusi pada pemahaman utama suatu teks, sehingga pemrosesan data menjadi lebih optimal. Proses ini sangat penting dalam analisis teks seperti klasifikasi dan analisis sentimen. Dengan data yang lebih terstruktur, tahap analisis berikutnya dapat dilakukan dengan lebih mudah dan efek.

### 3.4.4 Normalisasi

Normalisasi dalam tahap preprocessing adalah proses menyelaraskan teks dengan cara mengganti kata atau frasa tidak baku menjadi bentuk baku atau standar. Normalisasi digunakan untuk menyeragamkan istilah yang memiliki arti yang sama tetapi ejaannya berbeda, dapat disebabkan oleh kesalahan ejaan, singkatan kata atau “gaul”[21]. Hasil proses normalisasi dapat dilihat pada Tabel 7 berikut ini:

**Tabel 7.** Data Hasil Normalisasi

Proses	Hasil
Sebelum	@Lone_Lynx_ Tutup aj sekolah lah...KALO pemerintah gak mau kasih kurikulum yg jelas..itu tidur termasuk kurikulum merdeka belajar ya? Ya gak salah kalo masuk. Tidak ada perlindungan guru yg jelas. Masih ada guru honorer(harusnya pengajar standar gaji sama kecuali yg punya jabatan).
Sesudah	['lone', 'lynx', 'tutup', 'saja', 'sekolah', 'lah', 'kalau', 'pemerintah', 'tidak', 'mau', 'kasih', 'kurikulum', 'yang', 'jelasitu', 'tidur', 'termasuk', 'kurikulum', 'merdeka', 'belajar', 'ya', 'ya', 'tidak', 'salah', 'kalau', 'masuk', 'tidak', 'ada', 'perlindungan', 'guru', 'yang', 'jelas', 'masih',



Proses	Hasil
	'ada', 'guru', 'honorerharusnya', 'pengajar', 'standar', 'gaji', 'sama', 'kecuali', 'yang ', 'punya', 'jabatan']

Tabel 7 menunjukkan hasil dari proses normalisasi teks. Normalisasi adalah langkah dalam pemrosesan teks yang bertujuan untuk mengubah kata-kata tidak baku atau slang menjadi bentuk standar yang lebih umum. Dalam tahap ini, kata-kata seperti "gak" diubah menjadi "tidak", "kalo" menjadi "kalau", dan sebagainya. Hal ini penting untuk memastikan bahwa data yang dianalisis memiliki keseragaman dalam bentuk kata, sehingga hasil analisis lebih akurat.

Setelah proses normalisasi, teks menjadi lebih bersih dan seragam, yang memungkinkan pemrosesan lebih lanjut seperti stemming dan lemmatization dilakukan dengan lebih efektif. Normalisasi sangat bermanfaat dalam berbagai aplikasi NLP, termasuk analisis sentimen dan klasifikasi teks, karena mengurangi variasi kata yang tidak diperlukan dan meningkatkan konsistensi data.

### 3.4.5 Stemming

Proses dilanjutkan dengan menghilangkan imbuhan pada masing-masing kata sehingga menjadi kata dasar, dan dalam tahap ini bertujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat[22]. Hasil dari proses stemming dapat dilihat pada Tabel 8 berikut ini:

**Tabel 8.** Data Hasil Stemming

Proses	Hasil
Sebelum	@Lone_Lynx_ Tutup aj sekolah lah...KALO pemerintah gak mau kasih kurikulum yg jelas..itu tidur termasuk kurikulum merdeka belajar ya? Ya gak salah kalo masuk. Tidak ada perlindungan guru yg jelas. Masih ada guru honorer(harusnya pengajar standar gaji sama kecuali yg punya jabatan).
Sesudah	lone lynx tutup saja sekolah lah kalau perintah tidak mau kasih kurikulum yang jelasitu tidur masuk kurikulum merdeka ajar ya ya tidak salah kalau masuk tidak ada lindung guru yang jelas masih ada guru honorerharusnya ajar standar gaji

Tabel 8 menunjukkan hasil proses stemming yang bertujuan untuk mengubah kata-kata menjadi bentuk dasarnya. Proses ini dilakukan untuk menghilangkan imbuhan seperti awalan, sisipan, dan akhiran sehingga kata-kata yang memiliki makna sama dapat dianalisis dengan lebih efisien. Dengan demikian, teks menjadi lebih sederhana dan mudah diolah dalam tahap analisis selanjutnya.

Setelah proses stemming, teks mengalami perubahan signifikan dengan menghilangkan variasi kata yang tidak diperlukan. Kata-kata seperti "masuk", "melindungi", dan "pengajaran" dikonversi ke bentuk dasarnya, yakni "masuk", "lindung", dan "ajar". Hal ini memungkinkan model analisis teks bekerja dengan lebih optimal dalam mengenali pola dan tren dalam data, serta meningkatkan akurasi dalam pemrosesan bahasa alami.

## 3.4 Klasifikasi Model

Proses klasifikasi dalam penelitian ini dilakukan menggunakan dua metode pembelajaran mesin, yaitu Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Data yang telah melalui tahap preprocessing, seperti case folding, tokenisasi, stopword removal, normalisasi dan stemming, digunakan sebagai input. Data tersebut kemudian dibagi menjadi dua subset, yaitu data latih untuk melatih model dan data uji untuk menguji kinerja model.

### 3.5.1 Penerapan Algoritma SVM

SVM bekerja dengan mencari hiperplane optimal yang dapat memisahkan data berdasarkan kategori sentimen, yaitu positif, negatif, atau netral. Algoritma ini berfokus pada margin maksimal, yaitu jarak antara data terdekat dari setiap kelas terhadap hiperplane. Sistem pembelajaran Support Vector Machine (SVM) dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dan menerapkan learning biasa yang berasal dari teori pembelajaran statistik[23]. Hasil dari Klasifikasi dengan menggunakan algoritma SVM ditunjukkan pada Tabel 8 berikut ini.

**Tabel 8.** Hasil Klasifikasi Algoritma SVM

Sentimen	Accuracy	Precision	Recall	F1-Score
Negatif	91%	96%	41%	57%
Netral	91%	91%	100%	95%
Positif	91%	93%	17%	29%

Model SVM menunjukkan akurasi keseluruhan sebesar 91%, yang menandakan performa cukup baik dalam klasifikasi sentimen. Precision untuk kelas Negatif (0.96) dan Positif (0.93) menunjukkan model mampu mengidentifikasi kedua sentimen dengan baik. Namun, recall untuk kelas Negatif (0.41) dan Positif (0.17) masih rendah, menunjukkan kesulitan dalam menangkap semua contoh sentimen tersebut. Hal ini mengindikasikan bahwa model kurang optimal dalam mengenali variasi ekspresi sentimen negatif dan positif.

Sebaliknya, kelas Netral memiliki performa sangat baik dengan recall 1.00 dan precision 0.91, menandakan kemampuan model yang efektif dalam mengenali sentimen netral. F1-score menunjukkan performa beragam antar kelas, dengan nilai tinggi untuk kelas Netral (0.95), tetapi rendah untuk kelas Positif (0.29) dan Negatif (0.57). Hal ini menyoroti tantangan model dalam menangani ketidakseimbangan kelas, khususnya pada sentimen negatif dan positif. Meskipun demikian, SVM tetap menunjukkan akurasi yang baik secara keseluruhan. Hasil klasifikasi algoritma SVM ditampilkan pada tabel 8.

### 3.5.2 Penerapan Algoritma KNN

Algoritma K-Nearest Neighbors (KNN) juga digunakan untuk mengklasifikasikan sentimen masyarakat. Parameter k yang optimal ditentukan melalui uji coba beberapa nilai k. Hasil menunjukkan bahwa KNN dengan nilai  $k = 5$  memberikan akurasi tertinggi dibandingkan nilai lainnya. Dibandingkan dengan SVM, algoritma KNN cenderung lebih terpengaruh oleh ketidakseimbangan data. Hasil dari Klasifikasi dengan menggunakan algoritma KNN ditunjukkan pada Tabel 9 berikut ini:

**Tabel 9.** Hasil Klasifikasi Algoritma KNN

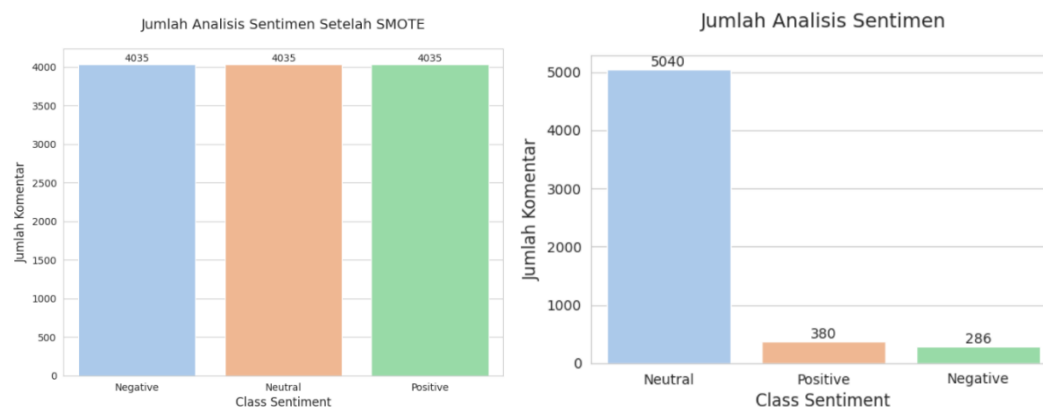
Sentimen	Accuracy	Precision	Recall	F1-Score
Negatif	89%	53%	18%	27%
Netral	89%	90%	99%	94%
Positif	89%	62%	16%	25%

Model KNN menunjukkan akurasi keseluruhan 89%, menandakan kemampuan yang cukup baik dalam klasifikasi sentimen. Precision untuk kelas Negatif (53%) dan Positif (62%) cukup moderat, tetapi lebih rendah dibandingkan precision untuk kelas Netral (90%). Hal ini menunjukkan keunggulan model dalam memprediksi sentimen netral. Namun, recall untuk kelas Negatif (18%) dan Positif (16%) sangat rendah, mengindikasikan kesulitan model dalam mendeteksi kedua sentimen tersebut.

Sebaliknya, recall untuk kelas Netral (99%) sangat tinggi, menandakan efektivitas model dalam mengenali sentimen netral. F1-score untuk kelas Netral (94%) sangat baik, tetapi rendah untuk kelas Negatif (27%) dan Positif (25%), mencerminkan performa yang tidak seimbang. Secara keseluruhan, meskipun akurasinya cukup baik, KNN lebih unggul dalam menangani sentimen netral tetapi kurang optimal dalam mendeteksi sentimen negatif dan positif. Hasil klasifikasi algoritma KNN ditampilkan pada tabel 9.

### 3.5 Evaluasi Model

Untuk mengatasi ketidakseimbangan data yang memengaruhi performa model, dilakukan evaluasi menggunakan teknik Synthetic Minority Oversampling Technique (SMOTE). Teknik ini menambah data pada kelas minoritas (positif dan negatif) untuk menyamakan distribusi dengan kelas mayoritas (netral). Dengan menambah sampel sintetis, SMOTE membantu model untuk lebih baik dalam menangani kelas minoritas tanpa mengorbankan informasi penting.



**Gambar 2.** Perbandingan Jumlah Analisis Sentimen Sebelum & Sesudah SMOTE

Gambar 2 menunjukkan perbandingan jumlah analisis sentimen sebelum dan sesudah penerapan teknik SMOTE. Sebelum SMOTE diterapkan, distribusi sentimen sangat tidak seimbang, dengan jumlah komentar pada kelas netral mencapai 5.040, jauh lebih banyak dibandingkan dengan kelas positif (380) dan negatif (286). Ketidakseimbangan ini dapat memengaruhi performa model, karena model cenderung lebih terfokus pada kelas yang memiliki jumlah data lebih besar (netral), sementara kelas positif dan negatif menjadi kurang terwakili dalam proses pelatihan.

Setelah SMOTE diterapkan, distribusi data menjadi seimbang di semua kelas, dengan jumlah komentar pada masing-masing kelas (negatif, netral, dan positif) meningkat menjadi 4.035. Penerapan SMOTE memperbaiki ketidakseimbangan ini dengan cara meningkatkan jumlah sampel pada kelas minoritas (positif dan negatif) melalui

teknik sintesis data. Hal ini memungkinkan model untuk belajar lebih baik dari setiap kelas secara merata, sehingga performa model dalam mengklasifikasikan sentimen menjadi lebih adil dan representatif terhadap keseluruhan dataset

**Tabel 11.** Tabel Hasil Evaluasi SVM Sesudah SMOTE

Sentimen	Accuracy	Precision	Recall	F1-Score
Negatif	94%	89%	71%	79%
Netral	94%	95%	100%	97%
Positif	94%	88%	43%	58%

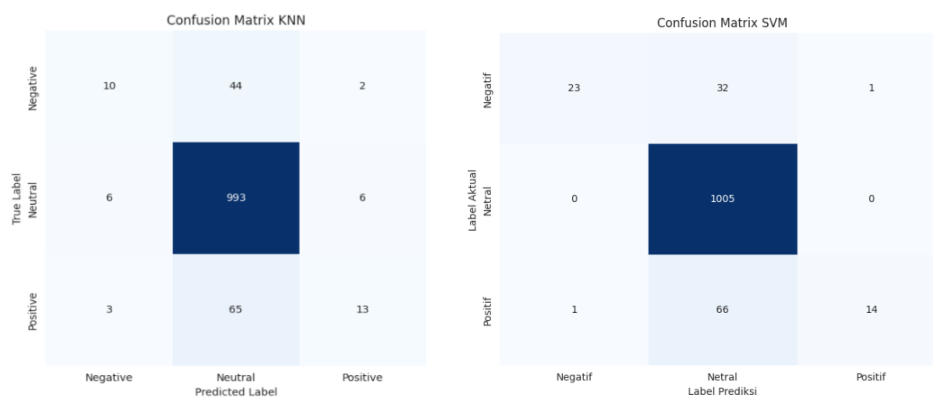
Setelah penerapan SMOTE, model SVM menunjukkan peningkatan akurasi signifikan hingga 94%. Precision untuk kelas Negatif (89%), Netral (95%), dan Positif (88%) meningkat, menunjukkan prediksi yang lebih akurat. Recall untuk kelas Negatif (71%) dan Netral (100%) juga mengalami perbaikan besar, khususnya pada kelas negatif yang sebelumnya rendah.

Namun, recall untuk kelas Positif (43%) masih relatif rendah, menandakan tantangan dalam mendeteksi sentimen positif. F1-score untuk kelas Negatif (79%) dan Netral (97%) tinggi, sementara kelas Positif (58%) tetap lebih rendah. Secara keseluruhan, penerapan SMOTE membantu meningkatkan keseimbangan model dalam menangani ketidakseimbangan kelas, meningkatkan deteksi sentimen negatif dan netral dengan lebih baik, namun masih ada ruang untuk peningkatan dalam mendeteksi sentimen positif secara lebih efektif. Hasil Evaluasi SVM sesudah SMOTE ditampilkan pada tabel 11.

**Tabel 12.** Tabel Hasil Evaluasi KNN Sesudah SMOTE

Sentimen	Accuracy	Precision	Recall	F1-Score
Negatif	31%	9%	77%	16%
Netral	31%	97%	26%	41%
Positif	31%	14%	67%	22%

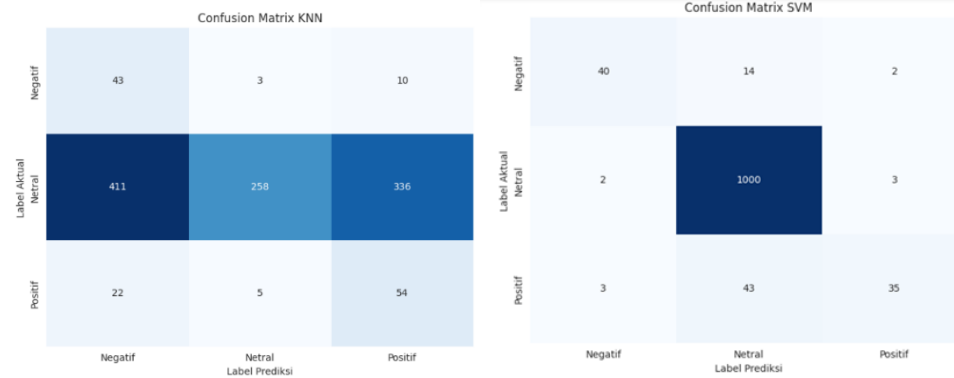
Setelah penerapan SMOTE, akurasi model KNN menurun signifikan menjadi 31%, menunjukkan bahwa meskipun data telah diseimbangkan, model ini masih kesulitan dalam melakukan klasifikasi sentimen secara menyeluruh. Precision untuk kelas Negatif (9%) dan Positif (14%) sangat rendah, sedangkan precision untuk kelas Netral (97%) menunjukkan kemampuan yang jauh lebih baik dalam memprediksi sentimen netral. Recall untuk kelas Negatif (77%) dan Positif (67%) tergolong cukup baik, namun recall untuk kelas Netral (26%) rendah, mengindikasikan keterbatasan model dalam mendeteksi sentimen netral secara konsisten. F1-score untuk kelas Negatif (16%) dan Positif (22%) sangat rendah, sementara kelas Netral (41%) juga kurang optimal, mencerminkan adanya ketidakseimbangan performa antara precision dan recall di semua kelas. Secara keseluruhan, meskipun SMOTE membantu menyeimbangkan data, model KNN masih menunjukkan kinerja buruk, terutama dalam mendeteksi sentimen negatif dan positif. Hasil Evaluasi KNN sesudah SMOTE ditampilkan pada tabel 12.



**Gambar 3.** Perbandingan Confusion Matrix SVM & KNN Sebelum SMOTE

Gambar 3 menunjukkan perbandingan confusion matrix untuk model SVM dan KNN sebelum penerapan teknik SMOTE. Pada confusion matrix SVM, model ini menunjukkan kinerja yang sangat baik dalam mengklasifikasikan kelas netral, dengan 1.005 prediksi benar dari 1.005 data aktual. Namun, model SVM mengalami kesulitan dalam mengklasifikasikan kelas negatif dan positif, dengan hanya 23 dari 56 data negatif dan 14 dari 81 data positif yang diklasifikasikan dengan benar, sementara sebagian besar data tersebut salah dikategorikan sebagai netral.

Sementara itu, confusion matrix untuk model KNN menunjukkan hasil serupa, dengan akurasi yang baik pada kelas netral (993 dari 1.005 prediksi benar), namun hanya 10 dari 56 data negatif dan 13 dari 81 data positif yang berhasil diklasifikasikan dengan benar. Sebagian besar data negatif dan positif pada model KNN juga salah diklasifikasikan sebagai netral. Kedua model menunjukkan kecenderungan kuat untuk mengklasifikasikan data sebagai netral, sementara kelas negatif dan positif kurang terdeteksi dengan baik, yang mencerminkan adanya ketidakseimbangan data yang signifikan sebelum penerapan teknik SMOTE.



Gambar 4. Perbandingan Confusion Matrix SVM & KNN Setelah SMOTE

Gambar 4 menunjukkan confusion matrix model SVM dan KNN setelah penerapan SMOTE untuk mengatasi ketidakseimbangan data. Pada SVM, performa meningkat pada kelas negatif dan positif, dengan 40 dari 56 data negatif dan 35 dari 81 data positif diklasifikasikan dengan benar. Namun, beberapa data negatif dan positif masih salah diklasifikasikan sebagai netral. Kelas netral tetap memiliki performa tinggi dengan 1.000 data benar. Sementara itu, confusion matrix KNN menunjukkan peningkatan yang lebih baik pada kelas minoritas, dengan 43 dari 56 data negatif dan 54 dari 81 data positif benar. Meski demikian, banyak data netral yang salah diprediksi sebagai negatif atau positif. Secara keseluruhan, SMOTE berhasil mengurangi bias kelas netral dan meningkatkan kemampuan model pada kelas minoritas.

### 3.6 Visualisasi

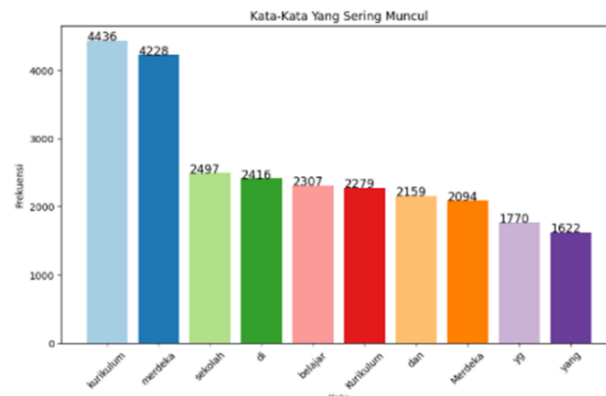
WordCloud menampilkan kata dengan frekuensi terbanyak, semakin banyak frekuensi kemunculan kata maka kata tersebut akan berukuran besar artinya kata tersebut memiliki frekuensi kemunculan kata yang lebih banyak pada dataset yang digunakan[24]. Semakin besar ukuran kata, semakin sering kata tersebut muncul dalam teks yang dianalisis. Wordcloud sering digunakan dalam analisis data teks untuk mengidentifikasi pola, tema utama, atau topik diskusi secara cepat. Dalam konteks penelitian ini, wordcloud digunakan untuk menggambarkan frekuensi kata dari komentar masyarakat tentang Kurikulum Merdeka



Gambar 5. WordCloud Sentimen

Gambar 5 merupakan WordCloud yang menampilkan berbagai kata yang sering muncul dalam komentar terkait Kurikulum Merdeka. WordCloud merupakan gambar yang menunjukkan daftar kata yang digunakan dalam teks[25]. Kata-kata yang paling dominan, seperti "kurikulum", "merdeka", dan "belajar", ditampilkan dengan ukuran huruf yang lebih besar, menunjukkan bahwa topik tersebut menjadi fokus utama dalam diskusi. Selain itu, kata-kata seperti "sekolah", "guru", "anak", dan "pembelajaran" juga muncul dengan cukup sering, menyoroti perhatian utama masyarakat terhadap implementasi kurikulum ini di lingkungan pendidikan.

Istilah-istilah lain seperti "Nadiem", "program", "kelas", dan "transformasi" menunjukkan keterkaitan diskusi dengan kebijakan pemerintah dan reformasi pendidikan. Pola kata dalam WordCloud ini juga mencerminkan sentimen yang beragam. Kata-kata seperti "tidak", "masih", dan "kenapa" mengindikasikan adanya kekhawatiran atau kritik terhadap implementasi Kurikulum Merdeka. Namun, kata-kata seperti "belajar", "anak", dan "pendidikan" menunjukkan harapan masyarakat terhadap peningkatan kualitas pendidikan. Secara keseluruhan, gambar ini memberikan wawasan tentang isu-isu yang paling sering dibahas dan sentimen yang mendasari diskusi publik tentang Kurikulum Merdeka.



**Gambar 4.** Kata Yang Sering Muncul

Gambar 4 diatas menunjukkan frekuensi kata-kata yang paling sering muncul dalam komentar masyarakat tentang Kurikulum Merdeka. Kata "kurikulum" dan "merdeka" mendominasi dengan frekuensi masing-masing 4436 dan 4228, yang menunjukkan bahwa kedua istilah ini menjadi inti utama dari diskusi masyarakat. Selanjutnya, kata-kata seperti "sekolah" (2497), "di" (2416), dan "belajar" (2307) mengindikasikan bahwa percakapan banyak berpusat pada implementasi Kurikulum Merdeka dalam lingkungan sekolah dan proses belajar mengajar.

Selain itu, kata "dan" (2159), "merdeka" (2094), "yg" (1770), dan "yang" (1622) menunjukkan pola penggunaan kata umum yang memperkaya konteks percakapan. Frekuensi tinggi kata-kata ini menggambarkan cakupan pembahasan yang luas, mulai dari kebijakan hingga pengalaman praktis di lapangan. Grafik ini memberikan wawasan penting tentang fokus dan perhatian masyarakat terhadap Kurikulum Merdeka, yang dapat membantu dalam mengevaluasi persepsi publik dan merancang strategi implementasi yang lebih efektif.

#### 4. KESIMPULAN

Penelitian ini telah membandingkan performa metode Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) dalam klasifikasi sentimen masyarakat terhadap implementasi Kurikulum Merdeka di sekolah di Indonesia. Hasil menunjukkan bahwa penerapan metode SMOTE berhasil mengatasi ketidakseimbangan data sentimen, sehingga menghasilkan distribusi kelas yang lebih merata dan meningkatkan performa model klasifikasi. Setelah SMOTE diterapkan, SVM menunjukkan akurasi yang tinggi dengan kemampuan yang stabil dalam mengenali kelas mayoritas (Neutral) dan kelas minoritas (Negative dan Positive). SVM mencapai akurasi 92%, dengan recall sempurna 100% untuk kelas Neutral, meskipun masih menghadapi tantangan pada kelas Negative (43%) dan Positive (40%). Di sisi lain, KNN menunjukkan performa keseluruhan yang lebih rendah dibandingkan SVM, dengan akurasi 31%. Meskipun demikian, KNN memiliki distribusi kesalahan yang lebih merata, meskipun kurang efektif dalam mengenali kelas minoritas. Penerapan SMOTE terbukti efektif dalam memperbaiki performa kedua model, namun SVM tetap unggul dalam stabilitas prediksi dan kemampuan menangani data yang tidak seimbang. Oleh karena itu, penelitian ini merekomendasikan penggunaan metode SVM untuk klasifikasi sentimen yang serupa, terutama pada dataset yang memiliki ketidakseimbangan kelas yang signifikan. Teknik penyalarsan data seperti SMOTE penting untuk meningkatkan kualitas prediksi model.

#### REFERENCES

- [1] K. Mahabatillah, "Analisis Pengembangan Kurikulum Merdeka dan Implementasinya," *Golden Age: Jurnal Pendidikan Anak Usia Dini*, vol. 8, no. 1, pp. 195–201, 2024, doi: 10.29313/ga.
- [2] L. Wanti and I. Chastanti, "Analysis of preparation in the independent curriculum implementation: Case study on IPAS learning," *BIO-INOVED : Jurnal Biologi-Inovasi Pendidikan*, vol. 5, no. 2, p. 250, 2023, doi: 10.20527/bino.v5i2.15493.
- [3] W. Darmawan, E. Jumiaty, and R. Sulistyaningsih, "Komparasi Metode Klasifikasi Untuk Analisis Sentimen Pengguna Twitter Terhadap Penerapan Kurikulum Merdeka," *IC-Tech*, vol. 18, no. 1, pp. 9–15, 2023, doi: 10.47775/icttech.v18i1.262.
- [4] W. Darmawan, M. Kurniawan Faizal, W. Setianto, and W. Hapsoro, "Analisis Sentimen Penerapan Kurikulum Merdeka Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor Dengan Forward Selection," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 12, no. 1, 2023, doi: 10.30591/smartcomp.v12i1.4634.
- [5] V. Krotov, L. Johnson, and L. Silva, "Tutorial: Legality and ethics of web scraping," *Communications of the Association for Information Systems*, vol. 47, no. 1, pp. 539–563, 2020, doi: 10.17705/1CAIS.04724.
- [6] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [7] M. Nanda Fahriza and N. Riza, "Analisis Sentimen Pada Ulasan Aplikasi Chat Generative Pre-Trained Transformer Gpt Menggunakan Metode Klasifikasi K-Nearest Neighbor(Knn)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 2, pp. 1351–1358, 2023, doi: 10.36040/jati.v7i2.6767.



- [8] A. Baita, Y. Pristyanto, and N. Cahyono, “Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (Svm) Dan K-Nearest Neighbor (Knn),” *Infos*, vol. 4, no. 2, pp. 42–42, 2021. <https://doi.org/10.24076/infosjournal.2021v4i2.687>
- [9] A. Budianto, R. Ariyuana, and D. Maryono, “Perbandingan K-Nearest Neighbor (Knn) Dan Support Vector Machine (Svm) Dalam Pengenalan Karakter Plat Kendaraan Bermotor,” *Jurnal Ilmiah Pendidikan Teknik dan Kejuruan*, vol. 11, no. 1, p. 27, 2019, doi: 10.20961/jiptek.v11i1.18018.
- [10] T. M. Permata Aulia, N. Arifin, and R. Mayasari, “Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinasi Covid-19,” *SINTECH (Science and Information Technology) Journal*, vol. 4, no. 2, pp. 139–145, 2021, doi: 10.31598/sintechjournal.v4i2.762.
- [11] D. Sartika, “Implementasi Algoritma K-Nearest Neighbour dalam Menganalisis Sentimen Terhadap Program Merdeka Belajar Kampus Merdeka (MBKM),” *Jurnal Buana Informatika*, vol. 14, no. 01, pp. 69–76, 2023, doi: 10.24002/jbi.v14i01.7178.
- [12] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, “Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 2, p. 121, 2021, doi: 10.22146/ijccs.65176.
- [13] T. A. Dewi and E. Mailoa, “Perbandingan Implementasi Metode Smote Pada Algoritma Support Vector Machine (Svm) Dalam Analisis Sentimen Opini Masyarakat Tentang Mixue,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 849–855, 2023, doi: 10.35870/jimik.v4i3.289.
- [14] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, “Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square,” *SisInfo : Jurnal Sistem Informasi dan Informatika*, vol. 5, no. 2, pp. 40–49, 2023, doi: 10.37278/sisinfo.v5i2.670.
- [15] A. W. Sari, T. I. Hermanto, and M. Defriani, “Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine,” *Sinkron*, vol. 8, no. 3, pp. 1366–1378, 2023, doi: 10.33395/sinkron.v8i3.12447.
- [16] S. Ernawati and R. Wati, “Evaluasi Performa Kernel SVM dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter dan VADER Lexicon,” *Jurnal Buana Informatika*, vol. 15, no. 01, pp. 40–49, 2024, doi: 10.24002/jbi.v15i1.7925.
- [17] D. Muhidin and A. Wibowo, “Perbandingan Kinerja Algoritma Support Vector Machine dan K-Nearest Neighbor Terhadap Analisis Sentimen Kebijakan New Normal,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 5, no. 2, p. 153, 2020, doi: 10.30998/string.v5i2.6715.
- [18] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, “Tweets Responding to the Indonesian Government’s Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel,” *Journal of Information Systems Engineering and Business Intelligence*, vol. 6, no. 2, p. 112, 2020, doi: 10.20473/jisebi.6.2.112-122.
- [19] A. D. Adhi Putra, “Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.
- [20] K. Pramayasa, I. M. D. Maysanjaya, and I. G. A. A. D. Indradewi, “Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE,” *SINTECH (Science and Information Technology) Journal*, vol. 6, no. 2, pp. 89–98, 2023, doi: 10.31598/sintechjournal.v6i2.1372.
- [21] M. H. Asnawi, I. Firmansyah, R. Novian, and R. S. Pontoh, “Perbandingan algoritma naïve bayes, k-nn, dan svm dalam pengklasifikasian sentimen media sosial,” *Prosiding Seminar Nasional Statistika*, vol. 10, no. June 2023, pp. 20–20, 2021, doi: 10.1234/pns.v10i.85.
- [22] N. Widya Utami and M. Artana, “Text Mining Dalam Analisis Sentimen Pembelajaran Daring Di Masa Pandemi Covid 19 Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Informatika Teknologi dan Sains*, vol. 4, no. 2, pp. 140–148, 2022, doi: 10.51401/jinteks.v4i2.2034.
- [23] E. Novianto, A. Hermawan, and D. Avianto, “Perbandingan Metode K-Nearest Neighbor dan Support Vector Machine Untuk Memprediksi Penerima Beasiswa Keringanan UKT,” *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, p. 654, 2024, doi: 10.30865/mib.v8i1.6913.
- [24] E. Suryati, Styawati, and A. A. Aldino, “Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM),” *Jurnal Teknologi Dan Sistem Informasi*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: <https://doi.org/10.33365/jtsi.v4i1.2445>
- [25] E. R. Lidinillah, T. Rohana, and A. R. Juwita, “Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine,” *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 154–164, 2023, doi: 10.37373/tekno.v10i2.440.