

Perbandingan Metode Naïve Bayes dan Support Vector Machine Pada Kasus Pembunuhan Vina Cirebon Berdasarkan Data X

Comparison of Naïve Bayes and Support Vector Machine Methods in the Vina Cirebon Murder Case Based on Data X

Salsa Desia Fitri*¹, Parjito²

¹Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

²D3 Sistem Informasi Akuntansi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

Email: ¹salsadesia03@gmail.com, ²djito@teknokrat.ac.id

*Penulis Koresponden

Received: 27 Desember 2024

Accepted: 04 Februari 2025

Published: 25 Februari 2025



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).
Copyright (c) 2025 JUSTINDO

ABSTRAK

Tindakan kriminal sering terjadi hampir setiap tahun. Seperti halnya pada tahun 2016 maupun di tahun sebelumnya atau sesudahnya banyak terjadi tindakan kriminal pembunuhan diberbagai kota yang semakin meningkat, contoh dari tindakan kriminal yaitu seperti tindakan curanmor, pencurian, kekerasan, pelecehan dan penipuan. Salah satu tindakan kriminal terjadi pada tahun 2016 yang artinya pembunuhan ini terjadi sejak 8 tahun yang lalu tetapi sampai saat ini kasus tersebut belum terungkap dikarenakan banyak nya kejanggalaan yang terjadi pada kasus tersebut, dan banyak nya muncul saksi-saksi baru yang muncul di kasus tersebut sehingga kasus tersebut sampai saat ini belum juga terungkap dalang di balik kasus pembunuhan ini. Kasus ini sedang banyak diperbincangkan oleh masyarakat di media sosial khususnya TikTok dan X, Karena dianggap sensasional, penekanan media pada isu ini telah memunculkan kekhawatiran serius, dan berpotensi mengeksploitasi. Oleh karena itu, opini masyarakat mengenai kasus pembunuhan ini tentu beragam. Oleh karena itu, analisis sentimen digunakan dalam penelitian ini untuk membandingkan metode Naïve Bayes dan Support Vector Machine untuk memastikan apakah opini publik terhadap suatu isu cenderung positif atau negatif berdasarkan data X. Menurut penelitian tersebut, 7146 titik data dikumpulkan antara bulan April dan Juli. Dengan tingkat akurasi 97% untuk Support Vector Machines serta 84,86% untuk Naïve Bayes, hasil klasifikasi model dari masing-masing teknik menghasilkan akurasi yang cukup akurat. Berdasarkan nilai akurasi dan evaluasi dari confusion matrix algoritma yang lebih optimal dalam analisis sentimen terkait kasus pembunuhan ini yaitu algoritma SVM dengan menghasilkan sentimen positif sebanyak 1384 data dan sentimen negatif sebanyak 1383 data.

Kata kunci: Analisis Sentimen; Naïve Bayes; Kasus Pembunuhan; Support Vector Machine

ABSTRACT

Criminal acts often occur almost every year. As in 2016 or in the previous or subsequent years, there were many criminal acts of murder in various cities that increased, examples of criminal acts such as motorcycle theft, theft, violence, harassment and fraud. One of the criminal acts occurred in 2016, which means that this murder occurred 8 years ago, but until now the case has not been solved due to the many irregularities that occurred in the case, and many new witnesses have emerged in the case so that the case has not yet revealed the mastermind behind this murder case. This case is being widely discussed by the public on social media, especially TikTok and X, the emphasis given by the media in this case is of sharp concern because it is considered sensational and has the potential to exploit. Therefore, public opinion regarding this murder case is certainly diverse. Therefore, sentiment analysis is used in this study to compare the Naïve Bayes and Support Vector Machine methods to determine whether public opinion on an issue tends to be positive or negative based on X data. According to the study, 7146 data points were collected between April and July. With a Naïve Bayes accuracy rate of 84.86% and a Support Vector Machine accuracy rate of 97%, the model

classification results from each technique produce fairly accurate accuracy. Based on the accuracy value and evaluation of the confusion matrix, the more optimal algorithm in sentiment analysis related to this murder case is the SVM algorithm by producing positive sentiments of 1384 data and negative sentiments of 1383 data.

Keywords: Naïve Bayes; Murder Case; Sentiment Analysis; Support Vector Machine

1. Pendahuluan

Tindakan kriminal sering terjadi hampir setiap tahun. Seperti halnya pada tahun 2016 maupun di tahun sebelumnya atau sesudahnya banyak terjadi tindakan kriminal pembunuhan diberbagai kota yang semakin meningkat, contoh dari tindakan kriminal yaitu seperti tindakan curanmor, pencurian, kekerasan, pelecehan dan penipuan. Salah satu tindakan kriminal terjadi pada tahun 2016 yang artinya pembunuhan ini terjadi sejak 8 tahun yang lalu tetapi sampai saat ini kasus tersebut belum terungkap (Suryasuciramdhan, Fitriany, et al., 2024).

Awalnya polisi mengira kasus ini adalah kecelakaan lalu lintas karna korban di temukan di jembatan layang sehingga tidak menimbulkan kecurigaan terhadap polisi. Tetapi setelah kasus ini di jadikan film dan pihak keluarga korban pun minta kepolisian untuk mengusut kembali kasus pembunuhan ini. Setelah kasus ini dikembangkan lagi oleh pihak kepolisian banyak nya saksi- saksi baru pada kasus ini. Dalam kasus pembunuhan ini Media masa semakin berperan penting dalam menyebarkan informasi dan membentuk pandangan masyarakat di tengah gelombang kemarahan dan kesedihan, tetapi yang dilakukan media dengan hal ini menjadikan perhatian masyarakat serta dianggap sensasional dan berpotensi mengeksploitasi. Kasus ini sedang banyak diperbincangkan di media sosial khususnya TikTok dan X, karena kasus tersebut belum diselesaikan selama kurang lebih delapan tahun, sehingga menimbulkan opini masyarakat terhadap kasus tersebut (Suryasuciramdhan, Dwi Mulyani, et al., 2024).

Opini publik terhadap kasus pembunuhan ini tentu beragam, ada opini publik yang resah karena khawatir kejadian ini akan terulang kembali, ada pula opini publik terhadap kasus yang belum tuntas, banyak pula masyarakat yang menginginkan agar pelaku pembunuhan ini segera ditangkap dan keluarga korban mendapatkan keadilan. Dengan memanfaatkan data tweet dari X, penelitian ini melakukan analisis sentimen terhadap kasus pembunuhan yang terjadi pada tahun 2016 .

Analisis sentimen, juga dikenal sebagai penambangan opini, adalah bidang luas pemrosesan bahasa alami, linguistik komputasional, dan penambangan teks yang bertujuan untuk mengevaluasi opini, pendapat, penilaian, sikap, penilaian, dan perasaan seseorang - baik penulis maupun pembicara - tentang suatu subjek, barang, layanan, perusahaan, atau orang, atau aktivitas tertentu (Nurrachmat Hidayat, 2023). Text mining adalah proses penggalian informasi berharga dari teks tak terstruktur menggunakan algoritma dan analisis bahasa alami untuk mengungkap pola dan informasi penting yang tersembunyi di dalam teks (Duei Putri, Nama and Sulistiono, 2022). Klasifikasi dalam analisis sentimen adalah proses pengelompokan dokumen atau teks ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang terkandung dalam teks (Kumala Sari and Randy Suryono, 2024). Metode klasifikasi ini sering digunakan untuk menganalisis teks dalam skala besar, seperti klasifikasi sentimen, klasifikasi topik, atau identifikasi spam dalam email (Setiawan and Suryono, 2024). Dalam analisis sentimen, algoritme klasifikasi seperti Naïve Bayes dan Support Vector Machine (SVM) sering digunakan untuk memprediksi kategori atau label yang tepat untuk setiap dokumen berdasarkan data tekstual (Singgalen, 2022).

Support Vector Machine (SVM), sebuah teknik pemisahan kelas untuk klasifikasi, memaksimalkan pemisahan antara sampel data terdekat setiap kelas dan hyperlane.dalam upaya untuk mengidentifikasi hyperplane pembeda terbaik di antara dua kelas (Sihombing, Hannie and Dermawan, 2021). Pendekatan serupa digunakan dalam penelitian sebelumnya untuk membandingkan Naïve Bayes dan Support Vector Machines (SVM). Teknik klasifikasi Naïve Bayes sederhana namun efisien dan sering kali memberikan hasil yang baik, terutama ketika ada lebih banyak fitur daripada sampel. Temuan dari survei ini, yang membahas tentang pelanggaran dan kebocoran data Twitter, menunjukkan bahwa masyarakat umum memiliki pendapat yang kurang

baik tentang keamanan dan pelanggaran data. Algoritma SVM digunakan untuk mengkategorikan (M. Anwar Sadat et al., 2024).

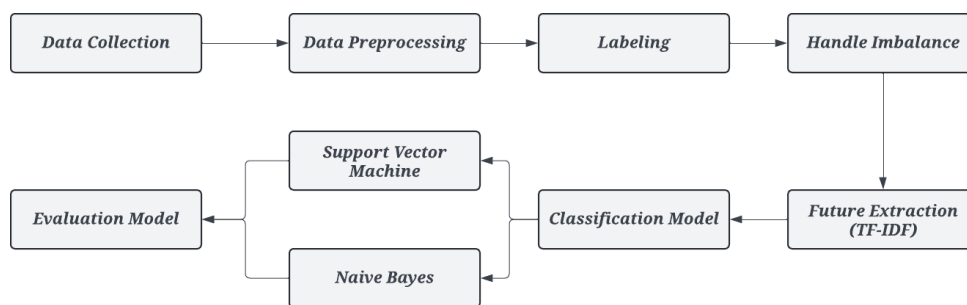
Meskipun Naïve Bayes secara numerik lebih unggul dalam proses klasifikasi dengan hasil yang dihitung, dengan Precision 97% dan Recall 97%, algoritma SVM digunakan dalam klasifikasi sentimen dengan Precision 80% dan Recall 93%, menunjukkan hasil yang akurat dan stabil dalam mengklasifikasikan tweet yang diambil dari Twitter (Zy and Hadikristanto, 2023). Penelitian lainnya yang membahas tentang analisis sentimen masyarakat terhadap isu penundaan pemilu 2024 di Twitter menggunakan teknik Support Vector Machine (SVM) dan Naïve Bayes, dari proses analisis sentiment yang dilakukan oleh penulis mendapatkan hasil perbandingan nilai akurasi metode SVM sebesar 91.61% dan naïve bayes 98.80% dari 585 data uji, terprediksi sebesar 429 data sebagai sentiment negatif dan 155 data sebagai sentiment positif. Untuk hasil prediksi sentiment negatif maka sebagian besar masyarakat tidak mempercayai berita hoax terhadap isu Penundaan Pemilu 2024 (Guru et al., 2024).

Penelitian yang dilakukan oleh Laisha Amilna Hayurian dan Nirwana Hendrastuty membandingkan Support Vector Machines (SVM) dan algoritma Naïve Bayes digunakan untuk menganalisis sentimen Twitter terkait boikot barang-barang Israel. Studi tersebut menemukan bahwa ketika 70% data dibagi menjadi pelatihan dan 30% menjadi pengujian, tingkat akurasi model algoritma SVM adalah 78%, sedangkan tingkat akurasi model algoritma naïve bayes adalah 84%. Model algoritma naïve bayes memiliki tingkat akurasi 84% dengan 80% data pelatihan dan 20% data pengujian, sedangkan model algoritma SVM memiliki tingkat akurasi 78%. mempertahankan tingkat akurasi SVM 84% dan tingkat akurasi tinggi 85%. Penulis menarik kesimpulan bahwa algoritma naïve bayes berkinerja lebih baik dari pada support vector machine berdasarkan hasil. (Hayurian and Hendrastuty, 2024). Dari temuan penelitian sebelumnya bisa dilihat dari perbandingan kedua metode yang memiliki akurasi tertinggi yaitu naïve bayes dengan permasalahan atau topik yang berbeda. Hasil penelitian tidak menggunakan optimasi SMOTE untuk mengatasi ketidakseimbangan data, oleh karena itu penelitian ini dilakukan untuk membandingkan kedua algoritma penggunaan optimasi SMOTE guna mengetahui tingkat akurasi mana yang lebih tinggi dengan topik yang berbeda.

Penelitian ini bertujuan untuk mengetahui opini masyarakat terhadap Kasus Pembunuhan yang terjadi pada tahun 2016 berdasarkan data X dan membandingkan algoritma Naïve Bayes dan Support Vector Machine (SVM) yang dikombinasikan dengan Synthetic Minority Oversampling Technique (SMOTE) untuk menangani masalah ketidakseimbangan data, serta berusaha mencari algoritma klasifikasi yang paling optimal dan membandingkan nilai akurasi, precision, recall dan F1-Score dari kedua algoritma tersebut. Penelitian ini juga dapat menjadi acuan masyarakat untuk mengetahui opini masyarakat terhadap kasus ini apakah cenderung positif atau negatif karena kasus pembunuhan ini sangat ramai diperbincangkan dimedia sosial khususnya media sosial X.

2. Metode Penelitian

Untuk mendapatkan informasi yang berkaitan dengan masalah penelitian, penelitian ini menggunakan X API dan kata kunci "kasus vira Cirebon" untuk mengumpulkan data opini publik mengenai kasus pembunuhan dua remaja, tahapan ini dinamakan data collection. Selanjutnya dilakukan preprocessing data, pelabelan data untuk mendapatkan sentimen positif dan negatif, Future Extraction menggunakan TF-IDF untuk mendapatkan data latih dan dan uji. Handle Imbalance dilakukan untuk menangani ketidakseimbangan data menggunakan optimasi SMOTE. Klasifikasi model menggunakan teknik Naïve Bayes dan Support Vector Machine (SVM). Naïve Bayes memiliki potensi yang baik untuk klasifikasi dalam hal akurasi dan perhitungan. Sedangkan SVM klasifikasi algoritma yang berbasis diskriminatif yang bermaksud membagi data menjadi dua kelas. Kedua metode ini sudah banyak digunakan untuk menemukan sentimen serta perbandingan antar kedua metode dengan kasus yang berbeda – beda. Yang terakhir evaluasi model untuk menilai akurasi dari suatu sistem klasifikasi. Skema penelitian dapat dilihat pada Gambar 1.



Gambar 1. Skema Penelitian

Gambar 1 menunjukkan skema penelitian, yang menjelaskan tentang proses penelitian ini dari awal hingga akhir yang memiliki 7 (tujuh) tahapan yaitu:

2.1. Data Collection

Data Collection atau pendekatan crawling digunakan untuk memperoleh data. Dengan menggunakan perangkat lunak komputer, metode crawling merupakan cara untuk mengambil data secara otomatis dari situs web. Dengan penggunaan teknik ini, data dapat diambil secara efektif dan sistematis dari berbagai sumber untuk digunakan dalam penelitian, analisis, dan pembuatan aplikasi, di antara aplikasi lainnya. (Algifari Rismawan *et al.*, 2023). Pada penelitian ini, dataset diperoleh melalui proses crawling pada platform media sosial X menggunakan bahasa pemrograman Python dengan library Harvest. Dataset tersebut berupa tweet pengguna mengenai kasus pembunuhan Vina Cirebon pada tahun 2016 menggunakan bahasa Indonesia dengan kata kunci “kasus Vina Cirebon”.

2.2. Data Preprocessing

Proses *preprocessing* data dilakukan karena dataset yang digunakan dari hasil *crawling* dari X, yang berupa teks tidak terstruktur. *Data Preprocessing* berisi beberapa tahapan antara lain, *case folding* yaitu merubah semua huruf menjadi kecil, data *cleaning* untuk hilangkan tanda baca seperti koma, tanda tanya, simbol, data duplikat, *Uniform Resource Locator* (URL), *tokenizing* digunakan untuk memecah kalimat menjadi kata, *stemming* untuk menghapus kata imbuhan dan mengubah menjadi kata dasar, sedangkan filtering untuk menghapus kata yang tidak memiliki makna khusus pada kalimat.

2.3. Labeling

Pelabelan adalah proses pemberian label atau klasifikasi pada data yang terkumpul. Pada tahap ini, data diberi label untuk menentukan apakah setiap kalimat dalam dataset memiliki makna positif, negatif, atau netral (Ilham and Pramusinto, 2023). Dalam penelitian ini, proses pelabelan data menggunakan pustaka *textblob* yang menghasilkan sentimen positif dan negative (Nofandi, Setiawan and Brata, 2023). Pelabelan sentimen negatif menghasilkan 4766 data dan sentimen positif 799 data.

2.4. Handle Imbalance

Dari hasil pelabelan yang diperoleh, terdapat ketidakseimbangan data, oleh karena itu perlu dilakukan proses Handle Imbalance menggunakan optimasi SMOTE. Salah satu metode untuk mengatasi ketidakseimbangan kelas pada data dalam machine learning adalah SMOTE (Hidayatullah and Umaidah, 2023). 80% dari data yang digunakan dalam penelitian ini adalah data pelatihan, sedangkan 20% adalah data pengujian.

2.5. Feature Extraction (TF-IDF)

Feature Extraction dilakukan Karena machine learning tidak dapat memahami kata-kata dan karakter dalam kumpulan data, ekstraksi selanjutnya dilakukan karena itu, agar data teks dapat dipahami, data tersebut harus disajikan sebagai angka. (Elisa and Rahman Isnain, 2024). Penelitian ini menggunakan Count-Vectorizer sebagai metode feature extraction. Sebelum dilakukan feature extraction, dataset perlu dibagi menjadi data latih dan data uji. Data pelatihan digunakan untuk

membuat model berdasarkan opini dari masing-masing tiga kategori: positif, negatif, dan netral. Untuk memilih model yang optimal, model yang dibangun diuji menggunakan data uji. (Yusliani *et al.*, 2023). Count-Vectorizer merupakan kelas fitur dari perhitungan numerik dan metode ekstraksi fitur teks yang umum digunakan yang mengubah teks menjadi matriks frekuensi kata (Rahayu *et al.*, 2022).

2.6. Classification Model

Pada penelitian ini *classification model* dilakukan menggunakan dua algoritma yaitu *Support Vector Machine (SVM)* dan *Naïve Bayes*. Pustaka *sklearn* dan bahasa pemrograman *Python* digunakan untuk membuat model klasifikasi. Kedua algoritma tersebut dijelaskan sebagai berikut:

2.6.1. Naïve Bayes

Naïve Bayes memiliki kemampuan kategorisasi yang kuat dalam hal komputasi dan akurasi. Karena merupakan vektor informasi objek yang membuat asumsi bahwa kualitas item ini bersifat independen, algoritma ini menentukan kemungkinan bahwa keputusan kategori keputusan itu benar. (Fachriza and Artikel, 2023). Bentuk umum teorema bayes ditunjukkan pada persamaan (1):

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)} \quad (1)$$

Nilai X pada persamaan (1) merupakan data yang kelasnya belum ditentukan, sedangkan Y merupakan kelas tertentu, untuk P (Y|X) merupakan hipotesis yang bergantung pada kondisi X, kemudian untuk P(X|Y) merupakan probabilitas sebelumnya dari suatu kelas berdasarkan kondisi hipotesis, P(Y) merupakan probabilitas dari hipotesis Y dan untuk P(X) merupakan probabilitas dari X (Tesalonika and Mailoa, 2024).

2.6.2. Support Vector Machine (SVM)

Support Vector machine adalah sistem kategorisasi diskriminatif yang menggunakan garis vektor yang dikenal sebagai hyperlane untuk membagi data ke dalam dua kelas (M. Anwar Sadat *et al.*, 2024). Hyperlane yang optimal pada linear kernel, dipresentasikan pada persamaan (2) dan (3):

$$(w.xi + b) \leq 1, yi = -1 \quad (2)$$

$$(w.xi + b) \leq 1, yi = -1 \quad (3)$$

Nilai xi merupakan data ke-i, nilai w.xi merupakan nilai bobot kelas data ke-i, b merupakan nilai bias dan yi merupakan kelas data ke-i.

2.7. Evaluation Model

Pada penelitian ini menggunakan metode *confusion matrix* sebagai evaluasi model. *Confusion matrix* merupakan sebuah teknik yang mudah dan efektif dalam mengukur kinerja sistem klasifikasi. Tujuan dari *confusion matrix* adalah untuk menilai akurasi dari suatu sistem klasifikasi data uji (Nurhidayat and Dewi, 2023).

3. Hasil dan Pembahasan

Penelitian ini menggunakan metode Naïve Bayes dan Support Vector Machine (SVM) untuk mengetahui opini publik terhadap kasus pembunuhan dua pemuda di Cirebon. Setelah itu, kedua metode tersebut dikontraskan untuk mengetahui metode mana yang lebih baik. Temuan dan pembahasan dari penelitian ini akan dijelaskan berikut ini:

3.1. Data Collection

Untuk menggunakan data dalam penelitian ini adalah opini atau komentar masyarakat mengenai kasus pembunuhan dua remaja Cirebon di media sosial X. Pengumpulan data menggunakan metode crawling dengan kata kunci “kasus Vina Cirebon” dalam bahasa Indonesia. Jumlah data yang berhasil dikumpulkan sebanyak 7146 data, data dikumpulkan sejak bulan April sampai dengan Juli 2024. Data yang telah terkumpul kemudian disimpan dalam format CSV. Hasil crawling data dapat dilihat pada Tabel 1.

Tabel 1. Crawling Data

Data Collection	
	Pengacara Ungkap Alasan Terpidana Kasus Pembunuhan Vina dan Eky Cirebon Ajukan Grasi Pengadilan Negeri (PN) Bandung kembali menggelar sidang praperadilan Pegi Setiawan tersangka utama dalam kasus pembunuhan Vina Cirebon. Sidang ini sebelumnya ditunda karena pihak termohon dari Polda Jawa Barat tidak hadir
@saya_rya	@secondlymine @ObiWan_Catnobi Tapi kayanya banyak juga ini anak petinggi polisi yg bermasalah. Kalo bener kabar pelaku pembunuhan Vina
@tvOneNews	Apakah benar Pegi Setiawan sebagai pelaku pembunuhan Vina dan Eky?
@saya_rya	@secondlymine @ObiWan_Catnobi Tapi kayanya banyak juga ini anak petinggi polisi yg bermasalah. Kalo bener kabar pelaku pembunuhan Vina

3.2. Data Preprocessing

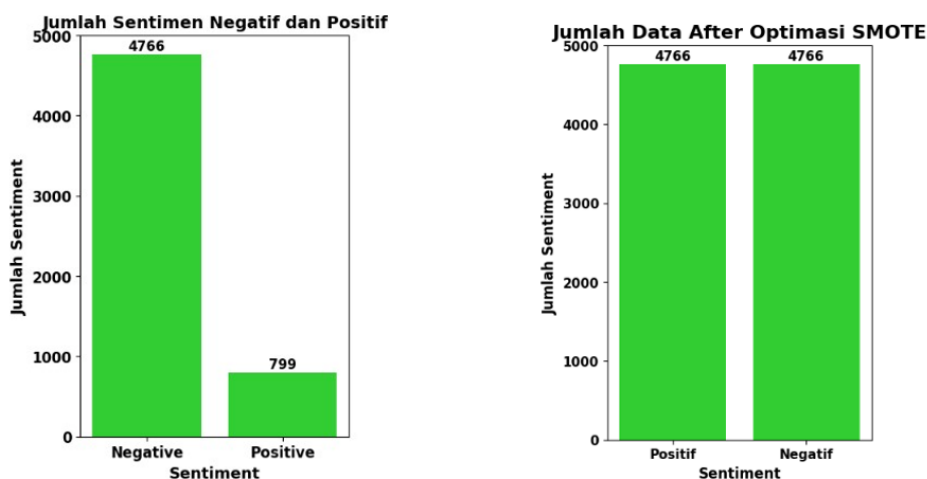
Data yang diperoleh merupakan data tidak terstruktur, oleh karena itu diperlukan preprocessing data. Preprocessing data memiliki beberapa tahapan, yaitu cleaning untuk menghilangkan tanda baca seperti koma, tanda tanya, simbol, data duplikat, Uniform Resource Locator (URL), Semua huruf diubah menjadi huruf kecil dengan melipat huruf. Kalimat dapat dipecah menjadi kata-kata individual menggunakan tokenisasi, istilah non-standar dapat diubah menjadi kata-kata standar menggunakan normalisasi, dan stemming. untuk menghilangkan imbuhan dan mengubahnya menjadi kata dasar, Tujuan penyaringan adalah untuk mengecualikan kata-kata dari kalimat yang tidak memiliki arti tertentu. Tabel 2 menampilkan hasil preprocessing.

Tabel 2. Hasil Preprocessing

Tahapan	Hasil
Data Tweet	@yusuf_dumdum Alhamdulillah dulu kasian pas dihadirin konferensi pers Setelah tak terbukti membunuh Vina Cirebon, Pegi Setiawan yang sebelumnya bungkam saat ditetapkan tersangka kini resmi bebas.
Data Cleaning	Alhamdulillah dulu kasian pas dihadirin konferensi pers Setelah tak terbukti membunuh Vina Cirebon, Pegi Setiawan yang sebelumnya bungkam saat ditetapkan tersangka kini resmi bebas.
Case Folding	alhamdulillah dulu kasian pas dihadirin pers Sempat bungkam sebelumnya, kini tersangka resmi dibebaskan setelah pembunuhan Vina Cirebon tak terbukti salah.
Tokenizing	"['dulu', 'kasian', 'pas', 'dihadirin', 'pers', 'dulu', 'dibungkam', 'saat', 'tersangka', 'kini', 'pegi', 'setiawan', 'resmi', 'bebas', 'usai', 'tak', 'terbukti', 'bunuh', 'vina', 'cirebon']"
Filtering	"['kasian', 'pas', 'dihadirin', 'pers', 'dibungkam', 'tersangka', 'pegi', 'setiawan', 'resmi', 'bebas', 'terbukti', 'bunuh', 'vina', 'cirebon']"
Stemming	kasi pas hadirin pers bungkam sangka pegi setiawan resmi bebas bukti bunuh vina cirebon

3.3. Labeling

Setelah dilakukan preprocessing data, data tersebut direduksi menjadi 5628 data dari data awal sebanyak 7146 data. Setelah mendapatkan data bersih, selanjutnya dilakukan pelabelan data menggunakan pustaka *TextBlob*. Dari total data tersebut, terdapat 4766 data yang tergolong sentimen negatif dan 799 data yang tergolong sentimen positif. Dengan hasil tersebut, mengalami ketidakseimbangan data antara sentimen negatif dan sentimen positif dapat menyebabkan model algoritma Naïve Bayes dan Support Vector Machine cenderung lebih terlatih untuk mengenali kata-kata negatif. Oleh karena itu, pada penelitian ini dilakukan optimasi menggunakan metode SMOTE untuk menyeimbangkan data sentimen negatif dan positif, dengan cara ini model tidak cenderung mendominasi satu sentimen tertentu. Hasil perbandingan klasifikasi dengan optimasi SMOTE dapat dilihat pada Gambar 2.



Gambar 2. Hasil Perbandingan Optimasi SMOTE

3.4. Feature Extraction (TF-IDF)

Pada tahap ini, pembobotan kata dilakukan dengan menggunakan pendekatan Count-Vectorizer untuk menentukan seberapa sering setiap kata muncul dalam setiap teks. Data dibagi menjadi 80% data pelatihan dan 20% data uji sebelum pembobotan kata. Pemodelan dilakukan menggunakan data pelatihan berdasarkan komentar positif dan negatif. Pembobotan kata dilakukan dengan menghitung frekuensi kemunculan kata dari setiap teks yang dipresentasikan dalam bentuk vector. Hasil ekstraksi fitur terdiri dari 4502 data sebagai data latih dan 1125 data sebagai data uji.

3.5. Classification Model

Setelah pelatihan model, prediksi data uji dibuat menggunakan model tersebut. Data pelatihan dari ekstraksi masa depan digunakan untuk melatih model Naïve Bayes dan Support Vector Machine.

3.5.1. Naïve Bayes

Hasil klasifikasi model Naïve Bayes menunjukkan persentase *fi-score* tertinggi (kelas tertinggi, 86%), persentase *recall* tertinggi (kelas positif, 96%), dan persentase *presisi* tertinggi (kelas negatif, 95%). Naïve Bayes memiliki tingkat akurasi keseluruhan sebesar 84,86%. Gambar 3 menampilkan hasil kategorisasi model Naïve Bayes.

Classification Report Naive Bayes				
	precision	recall	f1-score	support
negative	0.95	0.74	0.83	1440
positive	0.78	0.96	0.86	1420
accuracy			0.85	2860
macro avg	0.87	0.85	0.85	2860
weighted avg	0.87	0.85	0.85	2860

Gambar 3. Hasil Klasifikasi Model Naïve Bayes

3.5.2. Support Vector Machine (SVM)

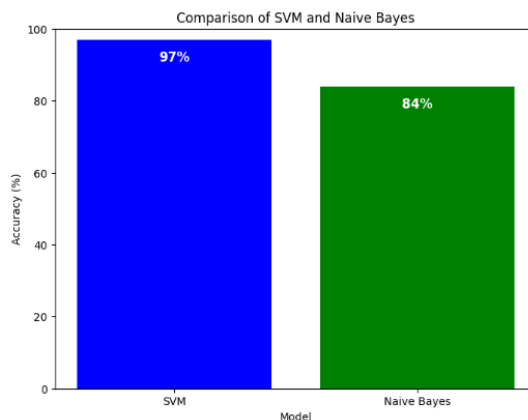
Hasil klasifikasi model SVM menunjukkan bahwa kelas negatif (97%) memiliki persentase *presisi* terbaik, dan kelas positif (97%) memiliki persentase *recall* tertinggi. Persentase *f1-score* juga sebesar 97%. SVM memiliki tingkat akurasi keseluruhan sebesar 97%. Gambar 4 menampilkan hasil klasifikasi model SVM.

Classification Report				
	precision	recall	f1-score	support
negative	0.97	0.96	0.97	1440
positive	0.96	0.97	0.97	1420
accuracy			0.97	2860
macro avg	0.97	0.97	0.97	2860
weighted avg	0.97	0.97	0.97	2860

Gambar 4. Hasil Klasifikasi Model SVM

3.6. Hasil Visualisasi Perbandingan Algoritma

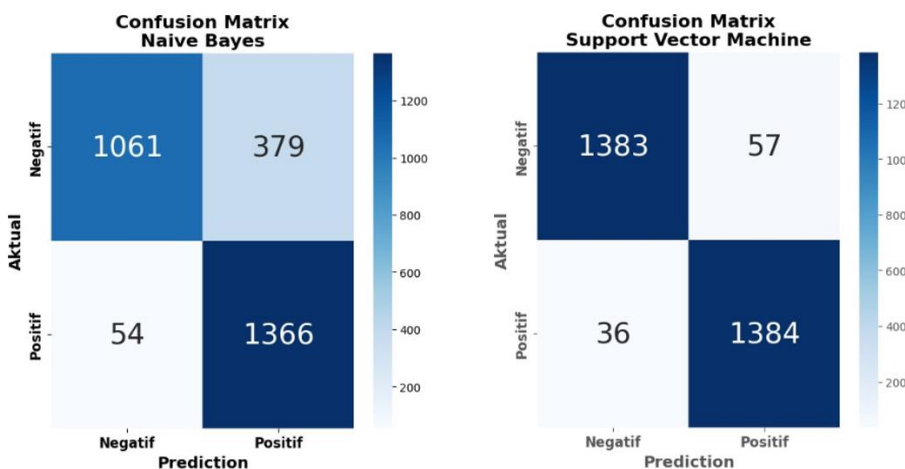
Visualisasi perbandingan perbandingan akurasi algoritma Naïve Bayes dan SVM. Algoritma Naïve Bayes mendapatkan akurasi sebesar 84%. Sedangkan algoritma SVM mendapatkan akurais sebesar 97%. Dari hasil akurasi tersebut dapat dilihat bahwa algoritma SVM memiliki akurasi lebih tinggi dari algoritma Naïve Bayes dalam analisis sentimen pada kasus pembunuhan seorang remaja tahun 2016. Sehingga algoritma SVM lebih efektif dalam analisis sentimen pada kasus penelitian ini. Hasil visualisasi perbandingan algoritma dalam dilihat pada Gambar 5.



Gambar 5. Visualisasi Perbandingan Algoritma

3.7. Evaluation Model

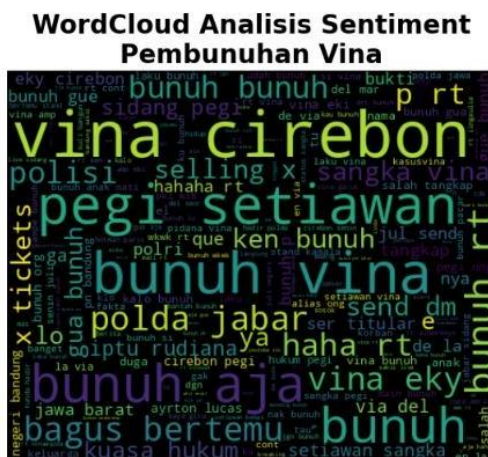
Evaluasi model dilakukan untuk mengukur nilai akurasi dan sistem klasifikasi data uji. Dalam penelitian ini, evaluasi model dilakukan dengan menggunakan metode matriks kebingungan untuk lebih memahami kinerja model dalam mengklasifikasikan sentimen yang memberikan gambaran menyeluruh tentang seberapa baik model dapat mengenali sentimen negatif dan positif.



Gambar 6. Confusion Matrix Naïve Bayes dan Support Vection Machine

Gambar 6 menunjukkan matriks konfusi kedua algoritma, kedua algoritma lebih dominan dalam memprediksi kata bernilai positif terkait kasus pembunuhan dua remaja Cirebon. Pada algoritma Naïve Bayes, nilai true negatif sebesar 1061 dan false negatif sebesar 379, sedangkan true positif sebesar 1366 dan false positif sebesar 54. Pada algoritma SVM, nilai true negatif sebesar 1383 dan false negatif sebesar 57, sedangkan true positif sebesar 1384 dan false positif sebesar 36.

Dalam penelitian ini, visualisasi wordcloud juga dilakukan untuk analisis data teks. Wordcloud merupakan representasi visual dari teks, kata-kata yang sering muncul diberi penekanan visual yang lebih besar dibandingkan dengan kata-kata yang jarang muncul. Hasil dari visualisasi wordcloud dapat dilihat pada Gambar 7.



Gambar 7. Visualisasi Wordcloud

4. Kesimpulan

Penelitian ini menunjukkan bagaimana penerapan metode Naïve Bayes dan Support Vector Machine (SVM), dipadukan dengan teknik SMOTE untuk menangani ketidakseimbangan data. Naïve Bayes memberikan solusi efisien dan cepat, sedangkan SVM menangani masalah klasifikasi lebih kompleks untuk menangani data dengan dimensi tinggi. Berdasarkan hasil penelitian, sentimen masyarakat terhadap kasus pembunuhan dua remaja Cirebon dapat diketahui dengan membandingkan pendekatan Naïve Bayes dan Support Vector Machine. Dengan hasil analisis sentimen menggunakan data dari X, hal tersebut dapat dilakukan dan mendapatkan hasil yang cukup bervariasi di kalangan masyarakat Indonesia dengan mayoritas memiliki opini atau pandangan positif terkait kasus pembunuhan ini dengan melihat matriks kebingungan yang ada. Hasil pengujian model Naïve Bayes memperoleh akurasi sebesar 84,86%, sedangkan pengujian model SVM memperoleh performa terbaik dengan akurasi sebesar 97%. Dengan demikian, dapat dikatakan bahwa algoritma SVM berkinerja lebih baik daripada Naïve Bayes dalam hal mengkategorikan tweet pada kasus pembunuhan ini. Saran untuk penelitian selanjutnya dapat menggunakan permasalahan baru dengan dataset yang lebih banyak dan tentunya dengan metode yang lebih bervariasi.

Daftar Pustaka

- Algifari Rismawan, S. et al. (2023) 'Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python', 10(3), pp. 167–178. Available at: <http://jurnal.mdp.ac.id>.
- Duei Putri, D., Nama, G.F. and Sulistiono, W.E. (2022) 'Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier', Jurnal Informatika dan Teknik Elektro Terapan, 10(1). Available at: <https://doi.org/10.23960/jitet.v10i1.2262>.
- Elisa, P. and Rahman Isnain, A. (2024) 'Comparison Of Random Forest, Support Vector Machine And Naive Bayes Algorithms To Analyze Sentiment Towards Mental Health Stigma', Jurnal Teknik Informatika (JUTIF), 5(1), pp. 321–329. Available at: <https://doi.org/10.52436/1.jutif.2024.5.1.1817>.

- Fachriza, M. and Artikel, H. (2023) 'url: <http://studentjournal.umpo.ac.id/index.php/komputek> Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest', *Komputek*, 7(2), pp. 49–58. Available at: <http://studentjournal.umpo.ac.id/index.php/komputek>.
- Guru, P. et al. (2024) 'Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine', *Jurnal Sains dan Teknologi*, 5(3), pp. 890–899. Available at: <https://doi.org/10.55338/saintek.v5i1.2789>.
- Hayurian, L.A. and Hendrastuty, N. (2024) 'Comparison Of Naive Bayes Algorithm And Support Vector Machine In Sentiment Analysis Of Boycott Israeli Products On Twitter', 5(3), pp. 731–738. Available at: <https://doi.org/10.52436/1.jutif.2024.5.3.1813>.
- Hidayatullah, H. and Umaidah, Y. (2023) Penerapan Naive Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt, *Jurnal Mahasiswa Teknik Informatika*.
- Ilham, A. and Pramusinto, W. (2023) '3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta', *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 2(2), pp. 539–547. Available at: <https://senafti.budiluhur.ac.id/index.php/senafti/index> (Accessed: 18 November 2024).
- Kumala Sari, P. and Randy Suryono, R. (2024) 'Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse', *Jurnal Mnemonic*, 7(1), pp. 31–39.
- M. Anwar Sadat et al. (2024) 'Comparison Of Algorithm Between Classification & Regression Trees And Support Vector Machine In Determining Student Acceptance In State Universities', *Jurnal Teknik Informatika (Jutif)*, 4(6), pp. 1589–1604. Available at: <https://doi.org/10.52436/1.jutif.2023.4.6.1565>.
- Nofandi, A., Setiawan, N.Y. and Brata, D.W. (2023) 'Analisis Sentimen Ulasan Pelanggan dengan Metode Support Vector Machine (SVM) untuk Peningkatan Kualitas Layanan pada Restoran Warung Wareg', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(1), pp. 458–466. Available at: <http://j-ptiik.ub.ac.id>.
- Nurhidayat, R. and Dewi, K.E. (2023) 'KOMPUTA: Jurnal Ilmiah Komputer dan Informatika Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek', 12(1). Available at: <https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-review?select=00.+Review.csv>.
- Nurrachmat Hidayat, F. (2023) 'Analisis Sentimen Masyarakat Terhadap Perekrutan PPPK Pada Twitter Dengan Metode Naive Bayes ... (Fajar Nurrachmat Hidayat, Sugiyono) Analisis Sentimen Masyarakat Terhadap Perekrutan PPPK Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine', *Jurnal Sains dan Teknologi*, 5(2), pp. 665–672. Available at: <https://doi.org/10.55338/saintek.v5i1.1359>.
- Rahayu, S. et al. (2022) 'Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP', *Edumatic: Jurnal Pendidikan Informatika*, 6(1), pp. 98–106. Available at: <https://doi.org/10.29408/edumatic.v6i1.5433>.
- Setiawan, A. and Suryono, R.R. (2024) 'Edumatic: Jurnal Pendidikan Informatika Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naive Bayes', 8(1), pp. 183–192. Available at: <https://doi.org/10.29408/edumatic.v8i1.25667>.
- Sihombing, L.O., Hannie, H. and Dermawan, B.A. (2021) 'Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naive Bayes Classifier', *Edumatic: Jurnal Pendidikan Informatika*, 5(2), pp. 233–242. Available at: <https://doi.org/10.29408/edumatic.v5i2.4089>.
- Singgalen, Y.A. (2022) 'Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM', *Building of Informatics, Technology and Science (BITS)*, 4(3). Available at: <https://doi.org/10.47065/bits.v4i3.2766>.
- Suryasuciramdhan, A., Dwi Mulyani, A., et al. (2024) 'Analisis Framing Film Vina: Sebelum 7 Hari dalam Media Sosial Tiktok dan X', 1(3), pp. 26–33. Available at: <https://doi.org/10.62383/filosofi.v1i3.136>.
- Suryasuciramdhan, A., Fitriany, R.M., et al. (2024) 'Analisis Framing Pemberitaan Film "Vina: Sebelum 7 Hari" Pada Media Online CNNIndonesia', *Jurnal Ilmiah Multidisiplin*, 1(5), pp. 154–159. Available at: <https://doi.org/10.62017/merdeka>.

- Tesalonika, R. and Mailoa, E. (2024) 'Implementasi Algoritma Naive Bayes Untuk Analisis Sentimen Isu Resesi Ekonomi 2023 Di Indonesia Pada Platform Twitter', *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 9(1), pp. 34–40. Available at: <https://doi.org/10.29100/jipi.v9i1.4288>.
- Yusliani, N. et al. (2023) Analisis Sentimen di Twitter Menggunakan Algoritma Artificial Neural Network.
- Zy, A.T. and Hadikristanto, W. (2023) 'Implementasi Algoritma Metode Naive Bayes dan Support Vector Machine Tentang Pembobolan dan Kebocoran Data di Twitter', *Bulletin of Information Technology (BIT)*, 4(1), pp. 49–56. Available at: <https://doi.org/10.47065/bit.v3i1>.