

Analisis Sentimen Komentar Youtube Terhadap Kasus Bjorka Dengan Membandingkan Navie Bayes dan SVM.

Rizky Haryanto Putro¹, Aria Hendrawan²

¹Universitas Semarang, Indonesia, e-mail: rizkyoppo030@gmail.com

²Universitas Semarang, Indonesia, e-mail: ariahendrawan@usm.ac.id

Info Artikel

Diajukan: 28-12-2025
Direvisi: 12-01-2026
Diterima: 25-01-2026
Diterbitkan: 31-01-2026

Kata Kunci:

Analisis Sentimen;
YouTube;
Bjorka;
Naïve Bayes;
Support Vector Machine.

Keywords:

Sentiment Analysis;
YouTube;
Bjorka;
Naïve Bayes;
Support Vector Machine.



Lisensi: cc-by-sa

Copyright © 2026 by Author.
Published by Faatuatua Media Karya

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terkait kasus kebocoran data yang melibatkan Bjorka dengan membandingkan kinerja algoritma Naive Bayes dan Support Vector Machine (SVM). YouTube, sebagai salah satu platform media sosial terbesar, telah menjadi forum bagi masyarakat untuk mengekspresikan opini dan persepsi mereka mengenai isu keamanan data dan privasi digital. Data penelitian terdiri dari komentar YouTube yang dikumpulkan dari beberapa video terkait kasus Bjorka selama periode 2023–2025. Data diproses melalui pra-pemrosesan teks, pelabelan sentimen menggunakan metode VADER Indonesia, penyeimbangan data untuk mengatasi ketidakseimbangan kelas, dan ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Selanjutnya, proses klasifikasi dilakukan menggunakan algoritma Naive Bayes dan Support Vector Machine. Hasil menunjukkan bahwa algoritma Support Vector Machine memiliki kinerja yang lebih baik dengan nilai akurasi 95,18% dan skor F1 95%, sedangkan algoritma Naive Bayes memperoleh nilai akurasi 73,81% dan skor F1 74%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma SVM lebih efektif dalam mengklasifikasikan sentimen komentar pengguna YouTube berbahasa Indonesia terkait kasus Bjorka.

Abstract

This study aims to analyze the sentiment of YouTube user comments related to the data breach case involving Bjorka by comparing the performance of the Naive Bayes algorithm and the Support Vector Machine (SVM) algorithm. YouTube, as one of the largest social media platforms, has become a forum for people to express their opinions and perceptions regarding data security and digital privacy issues. The research data consists of YouTube comments collected from several videos related to the Bjorka case during the period 2023–2025. The data was processed through text preprocessing, sentiment tagging using the Indonesian VADER method, data balancing to address class imbalance, and feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method. Next, the classification process was performed using the Naive Bayes algorithm and the Support Vector Machine. The results showed that the Support Vector Machine algorithm performed better with an accuracy of 95.18% and an F1 score of 95%, while the Naive Bayes algorithm achieved an accuracy of 73.81% and an F1 score of 74%. Based on these results, it can be concluded that the SVM algorithm is more effective in classifying the sentiment of Indonesian-language YouTube user comments related to the Bjorka case.

1. PENDAHULUAN

Perkembangan teknologi informasi memberikan kemudahan bagi masyarakat dalam memperoleh serta menyebarkan informasi melalui internet. Sejalan dengan perkembangan tersebut, platform media sosial seperti Facebook, Instagram, Twitter, dan YouTube mengalami

peningkatan jumlah pengguna yang signifikan. Media sosial tidak hanya berfungsi sebagai sarana hiburan, tetapi juga menjadi media utama bagi masyarakat dalam menyampaikan opini serta merespons berbagai isu publik.

YouTube merupakan salah satu platform berbagi video dengan tingkat keterlibatan pengguna yang tinggi, khususnya melalui kolom komentar. Berdasarkan laporan, lebih dari 95% pengguna internet mengakses YouTube, dengan lebih dari dua miliar pengguna aktif setiap bulan dan durasi tontonan mencapai miliaran jam setiap harinya[1]. Selain berfungsi sebagai media penyiaran, YouTube juga memungkinkan interaksi dua arah antar pengguna melalui kolom komentar. Komentar tersebut merepresentasikan opini publik secara langsung dalam jumlah besar, sehingga berpotensi menjadi sumber data yang bernilai untuk analisis sentimen. Analisis sentimen merupakan bagian dari text mining yang bertujuan untuk mengidentifikasi dan mengelompokkan opini pengguna ke dalam kategori tertentu, seperti positif, negatif, dan netral [2].

Salah satu isu yang memperoleh perhatian luas dari masyarakat Indonesia adalah kasus kebocoran data. Dalam konteks keamanan informasi, insiden kebocoran data telah terjadi secara berulang dan bahkan memunculkan anggapan bahwa Indonesia merupakan negara “open source” [3]. Isu kebocoran data kembali menjadi sorotan nasional pada tahun 2023 setelah munculnya peretas dengan nama samaran Bjorka yang diduga mempublikasikan data melalui situs *dataleaks*. Tindakan tersebut dinilai melanggar Undang-Undang Informasi dan Transaksi Elektronik (UU ITE), khususnya Pasal 30 terkait akses ilegal terhadap sistem elektronik, serta Undang-Undang Kependudukan mengenai penyebaran data pribadi seperti Nomor Induk Kependudukan (NIK) dan Kartu Keluarga (KK) secara tidak sah [4]. Bjorka juga mengklaim telah membocorkan data peserta BPJS Ketenagakerjaan. Isu ini kembali mencuat pada tahun 2025 seiring dengan pemberitaan mengenai proses penanganan kasus tersebut oleh aparat penegak hukum.

Banyaknya komentar publik pada video YouTube yang membahas kasus Bjorka menuntut adanya metode analisis yang mampu mengolah data teks secara efektif dan akurat. Penelitian dalam lima tahun terakhir menunjukkan bahwa algoritma Naive Bayes dan Support Vector Machine (SVM) merupakan metode yang umum digunakan dalam analisis sentimen teks berbahasa Indonesia. Naive Bayes dikenal memiliki keunggulan dari segi kesederhanaan model dan efisiensi komputasi, sedangkan SVM mampu memberikan tingkat akurasi yang tinggi dalam menangani data teks berskala besar [5]. Namun, masih terbatas penelitian yang mengombinasikan proses pelabelan sentimen menggunakan VADER Bahasa Indonesia dengan perbandingan kinerja kedua algoritma tersebut pada konteks isu kebocoran data di media sosial.

Berdasarkan permasalahan tersebut, penelitian ini menerapkan pendekatan *text mining* dan *machine learning* untuk melakukan analisis sentimen terhadap komentar YouTube yang berkaitan dengan kasus Bjorka. Penelitian ini berfokus pada perbandingan kinerja algoritma Naive Bayes dan Support Vector Machine dalam mengklasifikasikan sentimen opini publik. Hasil penelitian diharapkan dapat memberikan gambaran mengenai metode klasifikasi yang paling optimal untuk analisis sentimen komentar YouTube berbahasa Indonesia serta berkontribusi terhadap pengembangan teknologi informasi, khususnya dalam bidang *Natural Language Processing (NLP)*.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental, menerapkan teknik *Text mining* dan *machine learning* untuk melakukan analisis sentimen pada komentar pengguna YouTube terkait dengan pelanggaran kebocoran data yang dilakukan Bjorka. Pendekatan kuantitatif ini dipilih karena penelitian ini berfokus pada pemrosesan data numerik dari ekstraksi fitur teks dan mengevaluasi kinerja algoritma klasifikasi menggunakan metrik pengukuran spesifik.

Metode eksperimental digunakan untuk membandingkan kinerja dua algoritma klasifikasi, Naive Bayes dan Support Vector Machine (SVM), dalam mengklasifikasikan sentimen komentar publik ke dalam kategori positif, negatif, dan netral.

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini bersifat kuantitatif dengan pendekatan eksperimental. Kuantitatif yaitu metode yang masuk dalam bagian dari serangkaian investigasi sistematis terhadap fenomena dengan mengumpulkan data dan kemudian menghitungnya menggunakan statistik matematika atau komputasi [6] pada data komentar YouTube.

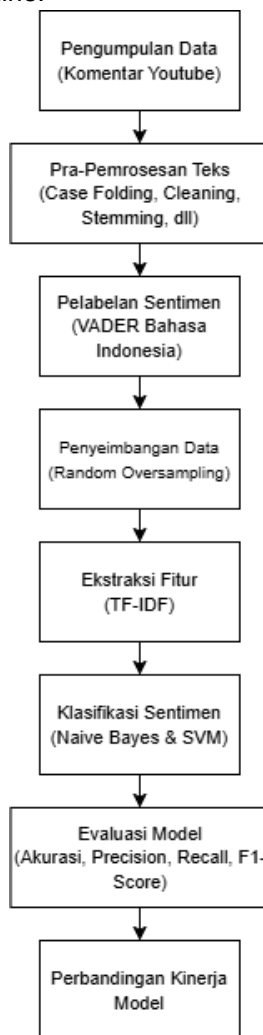
Ekstperimen dilakukan dengan melatih dan menguji dua algoritma, Naive Bayes dan Support Vector Machine (SVM), menggunakan data yang sama dan parameter evaluasi yang seragam. Hal ini memungkinkan perbandingan kinerja kedua algoritma secara objektif dan terukur.

2.2 Tahapan Penelitian

Tahapan penelitian disusun secara sistematis untuk memastikan bahwa pemrosesan data dan evaluasi model dapat dilakukan secara terstruktur dan terukur. Secara umum, tahapan penelitian meliputi:

1. Pengumpulan data komentar YouTube.
2. Pra-pemrosesan (preprocessing) teks.
3. Pelabelan sentimen menggunakan metode VADER Bahasa Indonesia.
4. Penyeimbangan data menggunakan teknik oversampling.
5. Ekstraksi fitur teks menggunakan metode TF-IDF.
6. Klasifikasi sentimen menggunakan algoritma Naive Bayes dan Support Vector Machine.
7. Evaluasi dan perbandingan kinerja model.

Untuk memperjelas alur tahapan penelitian yang dilakukan, diagram alur penelitian disajikan pada Gambar 1. Diagram tersebut menggambarkan urutan proses penelitian mulai dari pengumpulan data komentar YouTube hingga evaluasi dan perbandingan kinerja algoritma Naive Bayes dan Support Vector Machine.



Gambar 1 Diagram Alur Tahapan Penelitian Analisis Sentimen Komentar YouTube Kasus Bjorka

Tahapan penelitian dirancang secara sistematis untuk memastikan bahwa data komentar YouTube yang digunakan telah melalui proses pengumpulan, pra-pemrosesan, pelabelan sentimen, dan ekstraksi fitur yang optimal sebelum proses klasifikasi menggunakan algoritma Naive Bayes dan Support Vector Machine serta evaluasi kinerja model.

2.3 Jenis dan Sumber Data

Sebelum memasuki tahapan penelitian selanjutnya, terdapat dua aspek utama yang perlu dijelaskan, yaitu jenis data dan sumber data yang digunakan dalam penelitian ini.

1. Jenis data
Jenis data digunakan dalam penelitian ini adalah data teks berupa komentar pengguna YouTube yang mengandung opini dan sentimen terhadap kasus kebocoran data oleh Bjorka.
2. Sumber data
Sumber data diperoleh dari platform YouTube, khususnya kolom komentar pada beberapa video yang membahas kasus Bjorka pada periode 2023–2025.

2.4 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan teknik web scraping untuk mengambil komentar pengguna dari platform YouTube. Proses scraping memanfaatkan Antarmuka Pemrograman Aplikasi (API) YouTube atau pustaka pemrograman Python yang mendukung pengambilan data komentar berdasarkan ID video.

Teknik web scraping digunakan sebagai tahap awal penelitian untuk mendapatkan dataset komentar, yang kemudian disimpan dalam format CSV. Metode ini memungkinkan pengambilan data dalam jumlah besar secara otomatis dan efisien tanpa perlu menelusuri halaman web secara manual [7].

2.5 Pra-Pemrosesan Teks

Pra-pemrosesan teks merupakan tahap awal dalam penambangan teks dan pembelajaran mesin, yang bertujuan untuk mempersiapkan data mentah agar dapat dianalisis secara efektif atau digunakan dalam proses pemodelan. Tahap ini memainkan peran penting karena kualitas dan karakteristik data yang telah diproses akan secara langsung memengaruhi kinerja dan akurasi model klasifikasi yang sedang dibangun. Data teks yang tidak dibersihkan dengan benar berpotensi mengandung noise, yang dapat menurunkan kinerja model pembelajaran mesin yang digunakan [8].

Tahapan pra-pemrosesan teks yang dilakukan dalam penelitian ini meliputi:

1. Case folding.
2. Penghapusan URL, tag HTML, mention, dan hashtag.
3. Penghapusan emoji dan karakter non-alfanumerik.
4. Normalisasi kata tidak baku (*slang*).
5. Normalisasi frasa penting.
6. *Stopword removal*.
7. *Stemming* menggunakan algoritma Sastrawi.

2.6 Pelabelan Sentimen

Dalam analisis sentimen berbahasa Indonesia, beberapa studi telah mengusulkan pengembangan lexicon sentimen lokal untuk meningkatkan akurasi klasifikasi. Pengembangan ini dicapai dengan memperluas lexicon sentimen menggunakan istilah-istilah Indonesia dan frasa kontekstual spesifik, bersama dengan penyesuaian manual dan otomatis pada bobot valensi. Dengan pendekatan ini, metode analisis sentimen berbasis lexicon seperti VADER dapat lebih efektif merepresentasikan karakteristik linguistik dan konteks lokal tanpa mengubah rumus dasar untuk menghitung skor sentimen.

VADER (Valence Aware Dictionary and Sentiment Reasoner) adalah metode analisis sentimen yang bergantung pada lexicon dan aturan berbasis heuristik, yang dirancang khusus untuk menganalisis ekspresi sentimen dalam teks media sosial. Pendekatan ini menggunakan

kamus sentimen yang berisi skor valensi untuk setiap kata, memungkinkan sistem untuk secara otomatis mengidentifikasi dan mengukur kecenderungan emosional dalam teks secara otomatis [9].

Klasifikasi sentimen pada penelitian ini ditentukan berdasarkan nilai skor *compound* yang dihasilkan oleh metode VADER. Hasil pelabelan ini digunakan sebagai label kelas dalam proses pelatihan model klasifikasi. Ketentuan pengelompokan sentimen ditunjukkan pada Tabel 1 dibawah ini.

Tabel 1. Kategori Sentimen Berdasarkan Skor Compound VADER

<u>Rentang Skor Compound</u>	<u>Kategori Sentimen</u>
$\geq 0,05$	Positif
$-0,05 - 0,05$	Netral
$\leq -0,05$	Negatif

2.7 Penyeimbang Data

Distribusi kelas sentimen pada data komentar YouTube cenderung tidak seimbang, sehingga dapat memengaruhi kinerja model klasifikasi. Ketidakseimbangan kelas dapat menyebabkan model lebih dominan memprediksi kelas mayoritas. Oleh karena itu, penelitian ini menerapkan teknik *oversampling* untuk mengatasi ketidakseimbangan data dengan meningkatkan jumlah sampel pada kelas minoritas sehingga distribusi data antar kelas menjadi lebih seimbang [10].

Proses penyeimbangan data dilakukan hanya pada data pelatihan (*training data*) untuk menghindari terjadinya *data leakage* pada data pengujian. Proses penyeimbangan data hanya dilakukan pada data pelatihan untuk menghindari kebocoran data pada data pengujian. Dengan menerapkan *oversampling* pada data pelatihan, model dapat mempelajari karakteristik setiap kelas sentimen dengan cara yang lebih seimbang tanpa memengaruhi keaslian distribusi data pada data pengujian. Pendekatan ini bertujuan untuk menghasilkan evaluasi kinerja model yang lebih objektif dan mencerminkan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

2.8 Ekstraksi Fitur TF-IDF

Ekstraksi fitur teks dalam penelitian ini dilakukan dengan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), yang bertujuan untuk mengukur pentingnya suatu kata dalam sebuah dokumen relatif terhadap seluruh korpus. Metode TF-IDF mengubah data teks menjadi representasi vektor numerik dengan memberi bobot pada setiap kata berdasarkan frekuensi kemunculan dan kelangkaannya di seluruh dokumen.

Pendekatan ini efektif dalam merepresentasikan karakteristik data teks dan banyak digunakan dalam tugas klasifikasi sentimen karena dapat menyoroti kata-kata dengan makna yang signifikan dalam sebuah dokumen [11].

2.9 Algoritma Klasifikasi

Pada tahap klasifikasi, penelitian ini menggunakan dua algoritma *machine learning* yang umum digunakan dalam analisis sentimen: Naive Bayes dan Support Vector Machine (SVM). Kedua algoritma tersebut digunakan untuk mengklasifikasikan sentimen komentar pengguna YouTube ke dalam tiga kategori: positif, negatif, dan netral. Pemilihan kedua metode ini didasarkan pada karakteristik data teks dan efektivitasnya dalam menangani data berdimensi tinggi yang dihasilkan dari ekstraksi fitur TF-IDF.

1. Naïve Bayes

Naive Bayes adalah algoritma klasifikasi probabilistik berdasarkan Teorema Bayes, yang mengasumsikan independensi antar fitur. Algoritma ini banyak digunakan dalam analisis sentimen karena kemampuannya untuk memodelkan probabilitas suatu kata atau

frasa muncul dalam kategori sentimen tertentu, seperti positif, negatif, dan netral. Dengan menggunakan data pelatihan berlabel, model Naive Bayes dapat mempelajari pola hubungan antara kata-kata dalam teks dan kelas sentimen yang sesuai, sehingga memungkinkan klasifikasi yang efektif [12].

Studi ini menggunakan varian Complement Naive Bayes, yang dikembangkan untuk meningkatkan stabilitas kinerja pada data teks dengan distribusi kelas yang tidak seimbang. Varian ini telah terbukti lebih andal daripada Naive Bayes standar dalam mengatasi masalah ketidakseimbangan kelas yang umum ditemukan dalam data komentar media sosial.

2. Support Vector Machine(SVM)

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan membangun sebuah hyperplane optimal untuk memisahkan data ke dalam kelas-kelas tertentu. Prinsip utama SVM adalah memaksimalkan margin, yaitu jarak antara hyperplane pemisah dengan data terdekat dari masing-masing kelas, sehingga proses pemisahan antar kelas dapat dilakukan secara optimal dan akurat.

Dalam penelitian ini digunakan Linear Support Vector Machine, yang memproyeksikan data ke dalam ruang berdimensi tinggi dan memprediksi label kelas berdasarkan posisi data terhadap hyperplane pemisah yang dibentuk. Pendekatan linear dipilih karena sesuai untuk data teks hasil ekstraksi fitur menggunakan TF-IDF yang memiliki dimensi tinggi dan bersifat sparse. Dengan karakteristik tersebut, Linear SVM terbukti efektif dalam melakukan klasifikasi sentimen pada data teks berukuran besar [13].

2.10 Evaluasi Model

Performa model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Selain itu, confusion matriks digunakan untuk menganalisis distribusi kesalahan klasifikasi di setiap kelas sentimen. Untuk memastikan stabilitas model, evaluasi menggunakan metode *k-fold cross-validation*.

3. HASIL DAN ANALISIS

3.1 Hasil Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa komentar pengguna YouTube yang dikumpulkan dari beberapa video yang membahas kasus Bjorka pada periode 2023 hingga 2025. Dataset yang diperoleh mencerminkan opini publik terhadap isu kebocoran data di Indonesia. Berikut dibawah ini adalah gambar hasil dari pengumpulan data.

@Iskadarlskadar-y4o	Lupnama ,bukanitu di kepala sya,simpan seluru...	0
@globein1540	Seketika smua merasa tahu bjorkan 😂😂😂😂😂😂	0
@semutjantan5706	Udah ga gondrong bang	0
@fadlymooh7081	ketika gw denger pertama kali gw reflect bilan...	1
@dandyardika3909	junkyyy, keliatan bgt aura pemakai sabu hahha	0
@adinata_wjy	Dari pada uangnya di update utk keamanan, lebi...	0
@pujistory343		👍 0
@NurisFirdaus-fv7lj	Faktanya banyak orang yg tidak tau kalo bjorka...	1
@penikmatpekat5401	11:28 iya bener ketahuan saat mau beli mobil ...	0
@MUHAMMADRAHSYAFAHREZI-gj2rv		p 0
@nebulaharis	Bahas uang palsu yang ditemui di ATM	0
@anggakurniaputera9571	Om mau tanya gmna cara hubungi White Hack Om. ...	0

Gambar 2. Hasil Pengumpulan Data

3.2 Hasil Pra-Pemrosesan Data

Setelah dilakukan tahapan pra-pemrosesan, data komentar menjadi lebih bersih dan terstruktur. Penghapusan kata tidak relevan, normalisasi kata, serta proses *stemming* menghasilkan representasi teks yang lebih konsisten untuk proses klasifikasi. Berikut adalah hasil dari Pra-Pemrosesan Data.

Kalo hebat coba tangkap bjorka. Jangan cuma bac*t	kalau hebat coba tangkap bjorka jangan cuma bac
Materi podcast nya pinggir jurang	materi podcast nya pinggir jurang
Liat Pace komputer dah tentang heker hekeran pasti langsung tau aslinya.beritanya aja yang dilebi	liat pace komputer dah heker hekeran langsung tau asli berita lebih lebih
"Situs NASA" ya.bukan NASA nya.	situs nasa iya bukan nasa nya
Hoak..bocil cuma sites nasa kok ngaku2 sistem nasa hahahhaaa	hoak bocil cuma sites nasa kok ngaku2 sistem nasa hahahhaaa
Gw curiga sosok bejorka adalah putra aji sendiri	curiga sosok bejorka putra aji sendiri
ilmunya mantap bos qu 🙌🙌🙌🙌	ilmu mantap bos aku
Mantapp.. Amazing.. Pencerahan..	mantapp amazing cerah
bjorka buatan pemerintah buktinya hilang skrng. Hahaha	bjorka buat perintah bukti hilang skrng hahaha
Aku..beci bira. Aku di mayaya . Bika.	aku beci bira aku mayaya bika
Rerrrr Biroka bobol	rerrrr biroka bobol
Intinya gw tau bjorka itu siapa 😊	inti tau bjorka siapa

Gambar 3. Hasil Pra-pemrosesan

3.3 Hasil Pelabelan Sentimen

Pelabelan sentimen menggunakan metode VADER Bahasa Indonesia menghasilkan tiga kategori sentimen, yaitu positif, negatif, dan netral. Distribusi sentimen menunjukkan bahwa mayoritas komentar bersifat netral, diikuti oleh sentimen negatif dan positif. Berikut adalah gambar dari hasil pelabelan sentimen menggunakan metode VADER Bahasa Indonesia.

clean_text	vader_score	label
	0	netral
c fakta banyak orang tau kalau bjorka orang jabat korup	-8074	negatif
l 11 28 iya benar tahu mau beli mobil sport pake kartu kredit orang luar ceritain sama salah satu kalimantan emang keren betul sa	6124	positif
	0	netral
v bahas uang palsu ditemuin atm	0	netral
l paman mau tanya gmna cara hubung white hack paman soal papa kena tipu paman	-6124	negatif
i tuan takur jago nangkep maling ayan	0	netral
s cek dark web kak banyak aktor sana	0	netral
o pak dedy undang tifa dunk	0	netral
l coba bintang tamu nyari bjorka dh gausa aiueo	-3612	negatif
s 54 51 wa gb jenis pernah make bbrp jenis akhir make 2021 emang benar kalau orang delete message delete story tetep liat tanda	0	netral
o ketawa menit 37 27 37 33	0	netral
l kasus bjorka kasus dewi vatikan tuh	-3612	negatif
o orang polisi tangkap jahat lalu laku kejam interogasi lalu otak ambil jadi akal tangkap jahat	-9371	negatif
e sekarang aku bingung siapa nama dokter perempuan rs royal prima judes	0	netral
o anak punk apa	0	netral
l gimana masuk surga disangkain naksir	0	netral
o bjorka	-3612	negatif

Gambar 4 Pelabelan Sentimen menggunakan metode Vader Bahasa Indonesia

3.4 Hasil Klasifikasi Naive Bayes

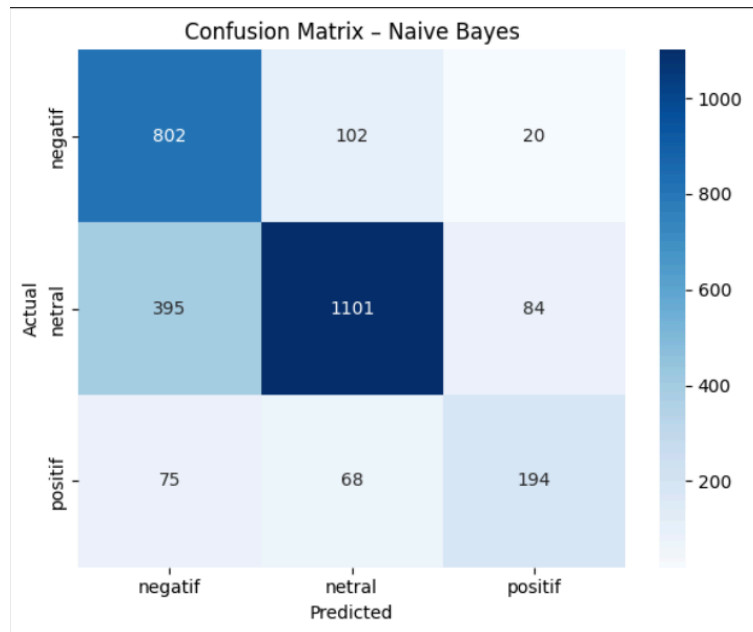
Hasil evaluasi kinerja algoritma Naive Bayes disajikan dalam Tabel 2, yang menampilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, visualisasi hasil evaluasi juga ditunjukkan pada Gambar 5, yang memberikan gambaran performa klasifikasi sentimen secara keseluruhan.

Tabel 2. Hasil Evaluasi Kinerja Algoritma Naive Bayes

Metrik evaluasi	Nilai
Accuracy	73,81%
Precision	0,76
Recall	0,74

F1-Score

0,74



Gambar 5 Confusion Matrix Naive Bayes

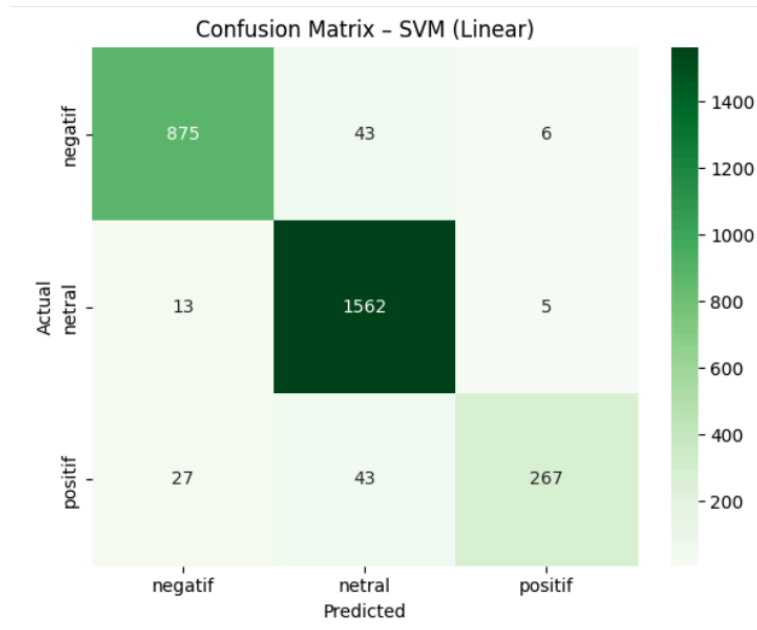
Seperti yang ditunjukkan dalam Tabel 2 dan gambar 5. Hasil pengujian algoritma Naive Bayes menunjukkan nilai akurasi sebesar 73,81% dengan skor F1 sebesar 73,96%. Model ini mampu mengklasifikasikan sentimen dengan cukup baik, namun masih mengalami kesalahan klasifikasi pada kelas tertentu.

3.5 Hasil Klasifikasi Support Vector Machine

Hasil evaluasi kinerja algoritma Support Vector Machine (SVM) disajikan pada Tabel 3, yang menampilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, visualisasi hasil klasifikasi sentimen menggunakan SVM ditunjukkan pada Gambar 6 untuk memberikan gambaran performa model secara lebih jelas.

Tabel 3. Hasil Evaluasi Kinerja Algoritma Naive Bayes

Metrik evaluasi	Nilai
Accuracy	95,18%
Precision	0,95
Recall	0,95
F1-Score	0,95



Gambar 6. Confusion Matrix SVM

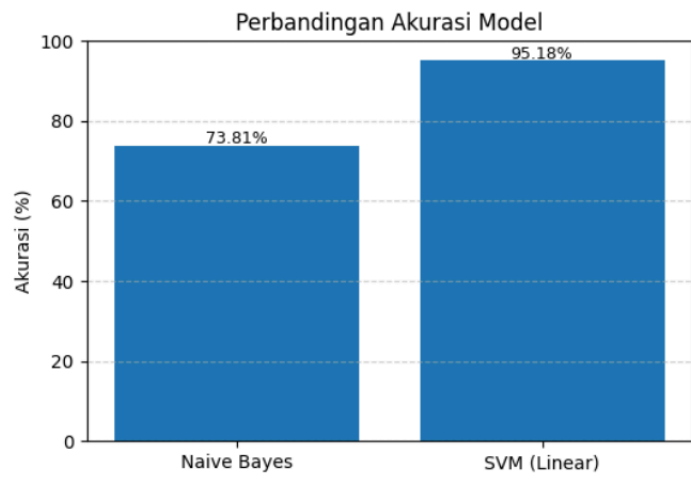
Berdasarkan hasil dari Tabel 3 dan Gambar 6, algoritma Support Vector Machine (SVM) menunjukkan performa yang lebih unggul dibandingkan Naive Bayes dengan nilai akurasi sebesar 95,18% dan skor F1 sebesar 95,07%. Hasil ini mengindikasikan bahwa SVM sangat efektif dalam menangani data teks hasil ekstraksi fitur TF-IDF yang berdimensi tinggi.

3.6 Perbandingan Kinerja Naive Bayes dan SVM

Untuk mengetahui algoritma terbaik dalam penelitian ini, dilakukan perbandingan kinerja antara algoritma Naive Bayes dan Support Vector Machine. Perbandingan kinerja antara algoritma Naive Bayes dan Support Vector Machine (SVM) ditunjukkan pada Tabel 4 dan Gambar 7.

Tabel 4 Tabel Perbandingan Naive Bayes dan SVM

<u>Algoritma</u>	<u>Accuracy</u>	<u>Precision</u>	<u>Recall</u>	<u>F-1 Score</u>
Naive Bayes	73,81%	0,76	0,74	0,74
SVM	95,18%	0,95	0,95	0,95



Gambar 7 Grafik Perbandingan Naive Bayes dan SVM

Berdasarkan hasil pengujian, algoritma SVM secara konsisten menghasilkan skor evaluasi yang lebih tinggi pada semua metrik dibandingkan dengan Naive Bayes. Hal ini menunjukkan bahwa SVM memiliki kemampuan generalisasi yang lebih baik dalam mengklasifikasikan sentimen komentar YouTube berbahasa Indonesia.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terkait kasus kebocoran data yang dikaitkan dengan Bjorka menggunakan pendekatan text mining dan machine learning. Proses penelitian dimulai dengan pengumpulan data komentar YouTube, diikuti dengan pra-pemrosesan teks untuk membersihkan dan menormalisasi data, pelabelan sentimen menggunakan metode VADER Indonesia, penyeimbangan data, dan ekstraksi fitur menggunakan metode TF-IDF.

Berdasarkan hasil pengujian dan evaluasi, model algoritma Naive Bayes dan Support Vector Machine sama-sama mampu mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Namun, Support Vector Machine menunjukkan kinerja yang lebih unggul dibandingkan dengan Naive Bayes. Hal ini menunjukkan bahwa Support Vector Machine lebih efektif dalam menangani data teks berdimensi tinggi yang diekstrak dari TF-IDF dan mampu memisahkan kelas sentimen secara lebih optimal.

Perbandingan kinerja kedua algoritma menunjukkan bahwa Support Vector Machine memberikan akurasi dan stabilitas klasifikasi yang lebih baik daripada Naive Bayes. Oleh karena itu, dapat disimpulkan bahwa Support Vector Machine adalah algoritma yang lebih tepat untuk analisis sentimen komentar YouTube terkait kasus kebocoran data Bjorka. Penelitian ini diharapkan dapat menjadi referensi untuk penelitian lebih lanjut dalam pengembangan analisis sentimen bahasa Indonesia, khususnya pada platform media sosial.

Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terkait kasus kebocoran data yang dikaitkan dengan Bjorka menggunakan pendekatan text mining dan machine learning. Proses penelitian dimulai dengan pengumpulan data komentar YouTube, diikuti dengan pra-pemrosesan teks untuk membersihkan dan menormalisasi data, pelabelan sentimen menggunakan metode VADER Indonesia, penyeimbangan data, dan ekstraksi fitur menggunakan metode TF-IDF.

Berdasarkan hasil pengujian dan evaluasi, model algoritma Naive Bayes dan Support Vector Machine sama-sama mampu mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Namun, Support Vector Machine menunjukkan kinerja yang lebih unggul dibandingkan dengan Naive Bayes. Hal ini menunjukkan bahwa Support Vector Machine lebih efektif dalam menangani data teks berdimensi tinggi yang diekstrak dari TF-IDF dan mampu memisahkan kelas sentimen secara lebih optimal.

Perbandingan kinerja kedua algoritma menunjukkan bahwa Support Vector Machine memberikan akurasi dan stabilitas klasifikasi yang lebih baik daripada Naive Bayes. Oleh karena itu, dapat disimpulkan bahwa Support Vector Machine adalah algoritma yang lebih tepat untuk analisis sentimen komentar YouTube terkait kasus kebocoran data Bjorka. Penelitian ini diharapkan dapat menjadi referensi untuk penelitian lebih lanjut dalam pengembangan analisis sentimen bahasa Indonesia, khususnya pada platform media sosial.

REFERENSI

- [1] M. Hudha, E. Supriyati, and T. Listyorini, "ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWAMENANTITERAWAN DENGAN METODE NAÏVE BAYES CLASSIFIER," *Jurnal Informatika dan Komputer) Akreditasi KEMENRISTEKDIKTI*, vol. 5, no. 1, pp. 2614–8897, 2022, doi: 10.33387/jiko.
- [2] Chely Aulia Misrun, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 207–215, Apr. 2023, doi: 10.37859/coscitech.v4i1.4790.
- [3] A. Turmudi Zy, A. Nugroho, A. Rivaldi, and I. Afriantoro, "Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes," *Jurnal Teknologi Informatika dan Komputer*, vol. 8, no. 2, pp. 202–213, Sep. 2022, doi: 10.37012/jtik.v8i2.1240.
- [4] Adhitya Karel Maulaya and Junadhi, "Analisis Sentimen Menggunakan Support Vector Machine Masyarakat Indonesia Di Twitter Terkait Bjorka," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 495–500, Dec. 2022, doi: 10.37859/coscitech.v3i3.4358.
- [5] J. Khatib Sulaiman, D. Atmajaya, A. Febrianti, H. Darwis, I. Artikel Abstrak, and K. Kunci, "Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 4, p. 2173, 2023.
- [6] V. Fitriyana *et al.*, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," 2023.
- [7] A. Sentimen, T. Komentar, V. Y. Menggunakan....., T. Muhayat, A. Fauzi, and D. J. Indra, "Analisis Sentimen Terhadap Komentar Video Youtube Menggunakan Support Vector Machines," 2023.
- [8] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [9] N. Fauziah, "Analisis Sentimen Publik Terhadap Kenaikan Tarif PPN di Indonesia dengan Pendekatan VADER," *Jurnal Akuntansi dan Keuangan*, vol. 12, no. 2, p. 228, Sep. 2024, doi: 10.29103/jak.v12i2.16796.
- [10] A. Simanungkalit, J. Panda, P. Naibaho, and A. De Kweldju, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Shopee Menggunakan Algoritma Naïve Bayes," 2024.
- [11] K. Tri Putra, M. Amin Hariyadi, and C. Crysdiyan, "PERBANDINGAN FEATURE EXTRACTION TF-IDF DAN BOW UNTUK ANALISIS SENTIMEN BERBASIS SVM," 2023.
- [12] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *Jurnal Ilmiah Computer Science*, vol. 2, no. 2, pp. 89–98, Jan. 2024, doi: 10.58602/jics.v2i2.24.
- [13] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.