

Sentiment Analysis of YouTube Comments on K-Pop Music Videos Using Naïve Bayes and SVM: A Case Study of Jung Jaehyun's 'Horizon'

Analisis Sentimen Komentar YouTube MV K-Pop Menggunakan Naïve Bayes dan SVM: Studi Kasus Jung Jaehyun 'Horizon'

Addriana Fatma Putri Indah Sari ¹⁾, Ade Eviyanti ²⁾, Ika Ratna Indra Astutik ³⁾, Novia Ariyanti ⁴⁾
^{1,2,3,4)} Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi : adeeviyanti@umsida.ac.id

Abstract. *This research aims to analyze the sentiment of YouTube comments on the music video "Horizon" by Jung Jaehyun by applying the Naïve Bayes and Support Vector Machine (SVM) algorithms. As a global phenomenon, K-pop presents both challenges and opportunities for understanding fan opinions through social media platforms, including YouTube, which provides public opinion data in the form of comments. The primary issue lies in the complexity of sentiment analysis on unstructured and imbalanced data, requiring an appropriate algorithmic approach. A total of 2,391 Indonesian language comments were collected using the YouTube API and processed through preprocessing stages such as data cleaning, tokenization, normalization, and the removal of common stopwords. The comments were manually labeled as positive or negative sentiments, then analyzed using Naïve Bayes, known for its simplicity, and compared with SVM equipped with a linear kernel. This research employs a quantitative method to evaluate the performance of both algorithms through metrics such as accuracy, precision, recall, and F1-score. The results indicate that SVM with a linear kernel achieved the highest accuracy of 98%, excelling in handling imbalanced data. However, Naïve Bayes provided competitive results with 97% accuracy, demonstrating efficiency in processing small datasets and ease of implementation. Both algorithms faced challenges in handling the negative class, as revealed by the confusion matrix. This study contributes to sentiment analysis methodologies by suggesting that Naïve Bayes is suitable for preliminary exploration, while SVM is more optimal for complex analyses. These findings are particularly relevant to the music industry in understanding fan sentiment as an indicator of a work's success.*

Keywords - sentiment analysis, naïve bayes, support vector machine, k-pop, youtube comments

Abstrak. *Penelitian ini bertujuan untuk menganalisis sentimen komentar YouTube pada video musik "Horizon" oleh Jung Jaehyun, dengan mengaplikasikan algoritma Naïve Bayes dan Support Vector Machine (SVM). Sebagai fenomena global, K-pop memunculkan tantangan dan peluang untuk memahami opini penggemar melalui media sosial, termasuk platform YouTube, yang menyediakan data opini publik dalam bentuk komentar. Masalah yang dihadapi adalah kompleksitas analisis sentimen pada data yang tidak terstruktur dan tidak seimbang, yang membutuhkan pendekatan algoritma yang tepat. Sebanyak 2.391 komentar berbahasa Indonesia dikumpulkan menggunakan YouTube API dan diproses melalui tahapan pra-pemrosesan, seperti pembersihan data, tokenisasi, normalisasi, dan penghapusan kata umum. Komentar tersebut dilabeli secara manual menjadi sentimen positif dan negatif, kemudian dianalisis menggunakan Naïve Bayes, yang dikenal karena kesederhanaannya, serta dibandingkan dengan SVM berkernel linier. Penelitian ini menerapkan metode kuantitatif untuk mengevaluasi performa kedua algoritma melalui metrik seperti akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa SVM dengan kernel linier mencapai akurasi tertinggi sebesar 98%, unggul dalam menangani data yang tidak seimbang. Namun, Naïve Bayes tetap memberikan hasil kompetitif dengan akurasi 97%, menunjukkan efisiensi dalam pemrosesan data kecil dan*

kemudahan implementasi. Keduanya memiliki kelemahan dalam menangani kelas negatif, seperti yang ditunjukkan oleh confusion matrix. Penelitian ini berkontribusi pada metode analisis sentimen dengan menyarankan bahwa Naïve Bayes cocok untuk eksplorasi awal, sementara SVM lebih optimal untuk analisis kompleks. Hasil ini relevan bagi industri musik untuk memahami sentimen penggemar sebagai indikator keberhasilan karya mereka.

Kata Kunci - analisis sentimen, naïve bayes, support vector machine, k-pop, komentar youtube

1. Pendahuluan

K-pop menjadi salah satu fenomena global yang menarik perhatian banyak pihak. Sebagai salah satu ekspor budaya terbesar Korea Selatan, K-pop berkontribusi pada fenomena Hallyu atau Korean Wave yang berhasil menyebarkan elemen budaya Korea Selatan, termasuk musik, fashion, dan drama, ke seluruh dunia. [1] Penggemar K-pop, yang dikenal sebagai "K-poppers," aktif menggunakan berbagai platform media sosial untuk mengakses dan mendukung karya idola mereka, salah satu di antaranya adalah YouTube.

YouTube adalah salah satu platform media sosial berbasis video yang populer, menyediakan konten dari berbagai genre. [2] Dan memungkinkan pengguna tidak hanya menikmati konten, tetapi youtube juga memiliki fitur umpan balik berupa suka, tidak suka, atau komentar terhadap video yang ditonton. [3] Hal ini menjadikan YouTube sebagai platform penting untuk mengumpulkan data opini publik, termasuk komentar dari penggemar K-pop. [2] Data ini, meskipun bermanfaat, memiliki tantangan signifikan dalam analisisnya, salah satunya adalah ketidakseimbangan data. Komentar positif cenderung mendominasi, sementara komentar negatif relatif sedikit, yang dapat memengaruhi akurasi hasil klasifikasi sentimen jika algoritma yang digunakan tidak dirancang untuk menangani data yang tidak seimbang. [4]

Sentiment Analysis, atau dikenal sebagai opini mining, merupakan kajian untuk menganalisis pendapat, emosi, evaluasi, sikap, dan sentimen publik terhadap berbagai objek seperti produk, layanan, individu, peristiwa, isu, dan topik tertentu. [5] Dalam perkembangannya, analisis sentimen dapat dilakukan secara manual melalui pengamatan langsung atau dengan bantuan teknologi seperti data crawling menggunakan API YouTube untuk mengumpulkan data komentar pengguna.

Dalam penelitian ini, mengangkat analisis sentimen yang merupakan cabang dari ilmu komputer yang bertujuan untuk mengelompokkan teks berupa komentar dari penggemar, dengan membagi komentar menjadi 2 kategori positif dan negatif. [6] Algoritma Naïve Bayes diimplementasikan untuk melakukan analisis sentimen pada komentar, dari penggemar K-pop pada video musik "Horizon" oleh Jung Jaehyun.

Algoritma Naïve Bayes dipilih karena kesederhanaan dan keefektifannya dalam klasifikasi data teks, serta kemampuannya dalam mengelompokkan data ke dalam kategori tertentu berdasarkan probabilitas. [2] Algoritma ini berfungsi untuk menganalisa suatu kalimat atau kata apakah kalimat itu bernilai positif atau bernilai negatif. [7] Selain menggunakan algoritma Naïve Bayes yang sederhana dan efektif, penelitian ini juga membandingkan performanya dengan metode Support Vector Machine (SVM), yang dikenal unggul dalam menangani klasifikasi data berdimensi tinggi, untuk mengevaluasi keunggulan masing-masing dalam klasifikasi sentimen positif dan negatif. [8] SVM memiliki kemampuan menangani data dengan dimensi tinggi serta memberikan hasil yang kompetitif dalam analisis sentimen dan klasifikasi data, meskipun pada dataset dengan ketidakseimbangan data. [9] Pemilihan kedua algoritma ini juga didasarkan pada keterbatasan algoritma lain, seperti kinerja yang kurang optimal pada algoritma berbasis keputusan pohon (decision tree) yang sering overfitting atau algoritma berbasis neural network yang membutuhkan lebih banyak data dan sumber daya komputasi.

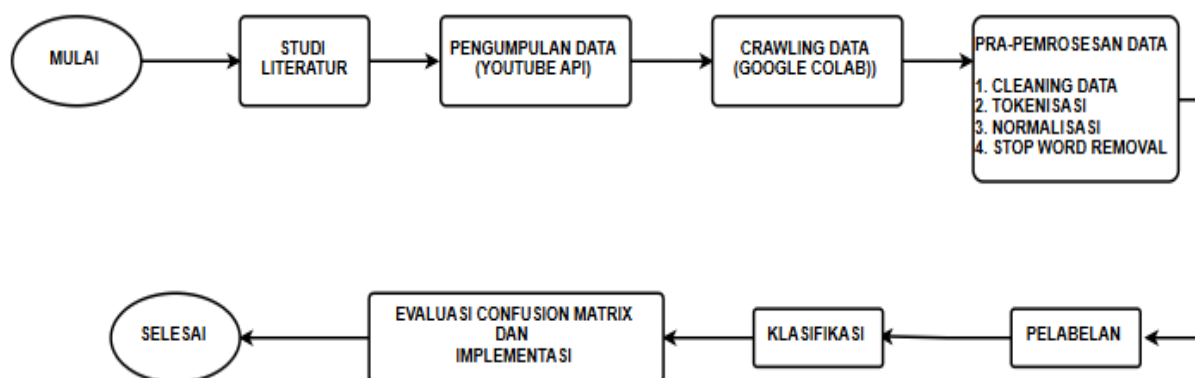
Pada penelitian sebelumnya yang berjudul “Analisis Sentimen Komentar YouTube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes”, yang dilakukan oleh [Firsttama et al. 2024], menunjukkan bahwa Naive Bayes dapat melakukan klasifikasi komentar dan mendapat nilai akurasi yang tinggi serta sangat relevan untuk memahami opini publik. [10] Gap penelitian terletak pada keterbatasan algoritma lain, seperti Decision Tree atau K-Nearest Neighbors (KNN), dalam menangani data komentar musik K-pop. Algoritma-algoritma ini sering kali memiliki performa yang kurang optimal pada dataset tidak terstruktur atau menghadapi kesulitan dalam menangani data dengan banyak dimensi, seperti teks komentar. Selain itu, penelitian sebelumnya cenderung lebih berfokus pada komentar produk atau layanan umum daripada karya seni seperti musik K-pop, yang memiliki karakteristik opini penggemar yang sangat spesifik dan sering kali emosional.

Dengan demikian, penelitian ini berusaha mengisi gap tersebut dengan mengevaluasi performa Naive Bayes dan SVM pada data komentar YouTube tentang musik K-pop. Penelitian ini diharapkan memberikan kontribusi baru dalam memahami bagaimana algoritma ini dapat dioptimalkan untuk memproses dan menginterpretasikan opini publik pada karya musik populer.

2. Metode Penelitian

Metode penelitian adalah serangkaian langkah atau prosedur yang diatur dengan sistematis untuk melaksanakan penelitian, dengan tujuan mendapatkan hasil yang diharapkan. [2] Bab ini akan menjabarkan metode yang akan diterapkan pada penelitian analisis sentimen komentar pada video musik K-POP di YouTube menggunakan algoritma Naive Bayes, dengan fokus pada studi kasus video musik "Horizon" oleh Jung Jaehyun.

Penelitian ini mengimplementasikan pendekatan kuantitatif dan menerapkan metode analisis sentimen. Pada gambar 1 dibawah ini merupakan hasil dari referensi jurnal Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naive Bayes. [2] Dimana tahapan tersebut ditambahkan lagi sesuai dengan alur struktur penelitian ini. Desain penelitian mencakup beberapa tahapan utama, sebagai berikut:



Gambar 1. Struktur Penelitian

Gambar 1, merupakan alur dari tujuh struktur penelitian yang akan di kerjakan oleh penulis, struktur diatas menunjukkan setiap proses identifikasi dan pemahaman masalah dalam penelitian yang dilakukan. Adapun proses dari struktur penelitian yaitu :

Proses pertama yaitu studi literatur untuk melakukan tinjau pustaka pada jurnal-jurnal terkait dengan analisis sentimen komentar agar penulis dapat memahami konsep-konsep yang relevan dengan topik yang akan diteliti, sumber-sumber teori seperti buku, artikel jurnal penelitian dari internet, yang berkaitan dengan analisis sentimen mengimplementasikan algoritma Naive Bayes sebagai acuan dalam penelitian. [11]

Proses kedua yaitu pengumpulan data yang terdiri dari komentar-komentar pengguna video musik “Horizon” di platform youtube, Pengumpulan data dilakukan melalui YouTube API (Application Programming Interface), yang memungkinkan akses ke statistik video dan informasi saluran YouTube. Proses ini dibangun dengan bahasa pemrograman Python pada notebook Google Colaboratory. [12] Dalam penelitian ini, YouTube API digunakan untuk mengakses dan mengunduh komentar-komentar tersebut. [13] Tahapan pengumpulan data meliputi mendaftar dan mendapatkan API Key dari YouTube dengan login ke Google Developer Console (<https://console.developers.google.com/>).

Proses ketiga yaitu crawling data, langkah ini diawali dari komentar dikumpulkan dengan cara melakukan crawling di media platform Youtube menggunakan Application Programming Interface (API youtube). [11] Pada tahap ini mengambil data komentar dari video musik K-pop di kanal youtube SM Town dengan judul “Horizon” oleh Jung Jaehyun. Tahap crawling data ini dilaksanakan pada 28 Mei 2024, menggunakan Google Colab sebagai alat untuk mengumpulkan data komentar.

Proses keempat adalah pra-pemrosesan data, proses ini dilakukan untuk membersihkan serta mempersiapkan data sebelum dianalisis. Adapun tahapan pra-pemrosesan data meliputi tahap pertama, cleaning data fungsinya untuk menghilangkan data-data yang mengandung tanda baca, emoticon dan numbering. [14] Tahap Kedua, tokenisasi yaitu memisahkan kalimat menjadi kata-kata atau token, proses tokenisasi juga dilakukan dengan menghapus tanda baca yang tidak relevan. [15] Tahap Ketiga, normalisasi yaitu mengganti kalimat ke bentuk dasar (stemming) dan mengkonversinya menjadi huruf kecil. Tahap Keempat, stop word removal : stop word removal digunakan untuk memilih kata-kata yang relevan dan menghilangkan kata-kata yang kurang penting. [16] Seperti, menghilangkan kata-kata umum yang tidak memberikan makna penting dalam analisis sentimen misalnya “dan”, “atau”, “dengan” dan sebagainya. Melalui proses-proses tersebut data yang semula tidak terstruktur akan melalui proses penyaringan dan pemrosesan sehingga siap digunakan untuk ke tahap berikutnya. [17]

Tabel 1. Hasil Pra-pemrosesan Data

publishedAt	authorDisplayName	textDisplay	stem
2024-09-03T13:05:41Z	@haura6480	Seger bet liat ni Cowo	seger bet liat ni cowo
2024-08-30T10:43:45Z	@octopus8978	â€™Aku masih disini	aku masih disini
2024-04-27T00:08:23Z	@refiana2303	Ganteng amat si jamaluddin###	ganteng amat si jamaluddin
2024-08-14T16:20:01Z	@user-vgid	PLASTIK AJA DIBANGGAIN****	plastik aja dibanggain
2024-10-08T18:37:46Z	@papakoji9239	ì¸ì¸ì¸,ì¸,œAku harap semakin banyak yang tau lagu iniê°• ë™ì• ì¼	aku harap semakin banyak yang tau lagu ini
2024-08-29T10:50:13Z	@naajae_14	â• ì¸, cinta banget sama Jaehyun • â• ì¸, • â• ì¸, •	cinta banget sama jaehyun
2023-08-21T16:01:55Z	@iom_nct127	ž'ë..., ì~í— ñf• í™œ ì• 'ë...,ëž~ë"ì œ¼ë©Sasaeng udah pada	sasaeng udah pada gila emang dari dullu penjarain aja

2023-08-10T10:58:56 Z	@zhangluobo2892	gila emang dari dullu, penjarain aja Pulang kerja dengerin Jaehyun langsung ilang capeknya	pulang kerja dengerin jaehyun langsung ilang capeknya
2023-08-08T09:18:03 Z	@finachyni3834	Jae kamu keren banget	jae kamu keren banget
2023-10-08T12:44:37 Z	@user_w1j4	SM gaa adil banget masa Jaehyun ga dikasih tour konser sendiri	sm gaa adil banget masa jaehyun ga dikasih tour konser sendiri

Proses kelima adalah pelabelan, pelabelan ini dilakukan secara manual dalam memberikan label sentimen positif dan negatif teks dokumen. [11] Selah dilakukannya pra-pemrosesan, data yang telah diperoleh akan melewati tahap pelabelan untuk mengidentifikasi sentimen, dimana pelabelan dilakukan secara manual oleh seorang ahli bahasa yang bernama Joko Susilo, S.Hum., M.Hum.[18] Pelabelan data ulasan dikategorikan sebagai positif atau negatif. Setelah itu, label positif bernilai 1 dan label negatif bernilai 0. [19] Setelah validasi manual, peneliti menerapkan undersampling untuk menyeimbangkan dataset, diikuti iterasi *k-fold cross-validation* untuk memastikan hasil yang general. Data dibagi menjadi beberapa *fold*, dengan tiap iterasi melibatkan satu *fold* sebagai data uji dan sisanya data pelatihan. Evaluasi menggunakan metrik akurasi, presisi, *recall*, dan F1-score, menghasilkan model yang lebih robust dalam menghadapi data tidak seimbang. Pada tabel 2 dibawah telah ditunjukkan contoh bentuk pelabelan manual yang akan digunakan dalam penelitian ini.

Tabel 2. Contoh pelabelan

Label	Text
Positif (1)	lagu ganteng dinyanyiin orang ganteng
Negatif (0)	ini produk boikot gila

Proses keenam merupakan proses klasifikasi algoritma, proses klasifikasi akan diterapkan dengan menggunakan algoritma Naive Bayes. Setelah data melewati tahap pelabelan, data tersebut akan diproses dengan metode Naive Bayes. Berikutnya mesin akan dengan mudah memprediksi pola dan mengklasifikasikannya pada dua bagian (positif dan negatif). [2] Metode ini hanya menggunakan jumlah data berelatif kecil dalam menentukan estimasi parameternya, sehingga sering digunakan dalam klasifikasi teks, seperti analisis sentimen komentar. Rumus 1 dibawah ini merupakan rumus Naive Bayes:

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (1)$$

Dimana dijelaskan bahwa rumus 1 sebagai berikut:

$P(A|B)$: Probabilitas hipotesis A benar berdasarkan bukti B (probabilitas posterior).

$P(B|A)$: Probabilitas memperoleh bukti B jika hipotesis A benar (likelihood).

$P(A)$: Probabilitas awal hipotesis A (prior).

$P(B)$: Probabilitas memperoleh bukti B (evidence).

Dan proses ketujuh adalah evaluasi confusion matrix dan implementasi, dimana pada proses ini akan dilakukan analisis hasil klasifikasi sentimen berdasarkan perbandingan antara hasil aktual dan prediksi. Perhitungan menggunakan Confusion Matrix dapat menunjukkan sentimen positif dan negatif untuk kedua kelas (positif dan negatif). [20] tabel 3 dibawah ini adalah contoh tabel rumus confusion matrix:

Tabel 3. Rumus Evaluasi Confusion matrix

	Prediksi Positif	Prediksi Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan pada tabel 3 dan rumus diatas:

TP (True Positive) : Data positif yang teridentifikasi dengan benar.

TN (True Negative) : Data negatif yang teridentifikasi dengan benar.

FP (False Positive) : Data negatif yang teridentifikasi sebagai positif.

FN (False Negative) : Data positif yang teridentifikasi sebagai negatif.

Pada rumus 2, 3, 4 dan 5 dibawah ini merupakan rumus dari matrix evaluasi yang akan digunakan dalam menentukan hasil akurasi dari metode Naïve Bayes.

$$Accuracy = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = 2 * \frac{Precision \cdot Recall}{Precision+Recall} \quad (5)$$

Implementasi penelitian menggunakan pemrograman Python pada Google Colaboratory.[21] Penelitian ini menguji kinerja algoritma Naïve Bayes dan membandingkannya dengan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen komentar pada video musik K-pop di YouTube, untuk menentukan model yang paling efektif dalam klasifikasi sentimen positif dan negatif.

3. Hasil

Penelitian ini menganalisis sentimen komentar pada video musik K-pop "Horizon" oleh Jung Jaehyun menggunakan algoritma Naïve Bayes. Peneliti mengumpulkan data komentar awal sejumlah 13.000 lebih melalui Youtube API, namun setelah melewati beberapa tahapan pada pra-pemrosesan data dan juga komentar dipilah dan dipilih yang hanya berbahasa Indonesia. Sehingga data komentar berbahasa Indonesia yang di dapat berjumlah 2.391 komentar.

Tabel 4. Sampel Data

publishedAt	authorDisplayName	stem
2024-09-03T13:05:41Z	@haura6480	seger bet liat ni cowo
2024-08-30T10:43:45Z	@octopus8978	aku masih ada disini

2024-04-27T00:08:23Z	@refiana2303	ganteng amat si jamaluddin
2024-08-14T16:20:01Z	@user-vgid	plastik aja dibanggain
2024-10-08T18:37:46Z	@papakoji9239	aku harap semakin banyak yang tau lagu ini
2024-08-29T10:15:13Z	@naajae_14	cinta banget sama Jaehyun
2023-08-21T16:01:55Z	@iom_nct127	sasaeng udah pada gila emang dari dullu penjarain aja
2023-08-10T10:58:56Z	@zhangluobo2892	pulang kerja dengerin jaehyun langsung ilang capeknya
2023-08-08T09:18:03Z	@finachyni3834	jae kamu keren banget
2023-10-08T12:44:37Z	@user_w1j4	sm gaa adil banget masa jaehyun ga dikasih tour konser sendiri

Pada tabel 4 diatas merupakan sampel data awal komentar yang akan melalui proses pelabelan, dimana data akan dibagi menjadi dua kategori positif dan negatif, dari proses pelabelan tersebut menghasilkan data label positif (1) dan data label negatif (0) seperti pada tabel 5, dibawah ini.

Tabel 5. Pelabelan

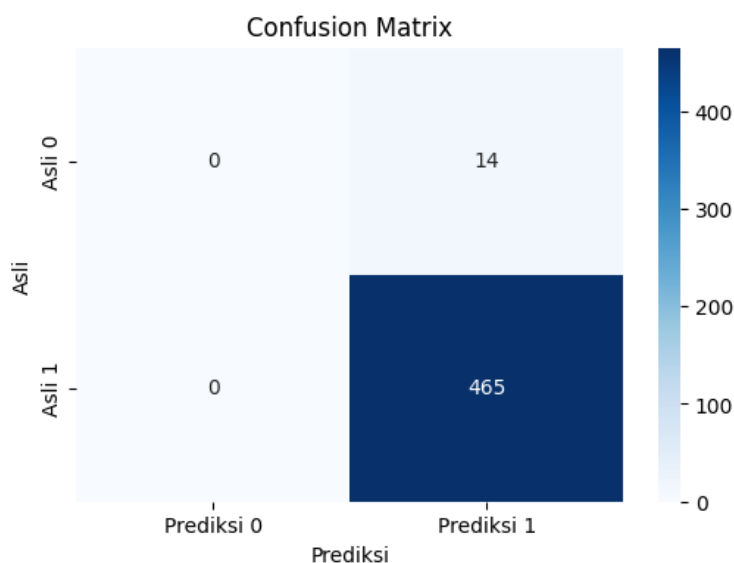
publishedAt	authorDisplayName	stem	Label
2024-09-03T13:05:41Z	@haura6480	seger bet liat ni cowo	1
2024-08-30T10:43:45Z	@octopus8978	aku masih ada disini	1
2024-04-27T00:08:23Z	@refiana2303	ganteng amat si jamaluddin	1
2024-08-14T16:20:01Z	@user-vgid	plastik aja dibanggain	0
2024-10-08T18:37:46Z	@papakoji9239	aku harap semakin banyak yang tau lagu ini	1
2024-08-29T10:15:13Z	@naajae_14	cinta banget sama Jaehyun	1
2023-08-21T16:01:55Z	@iom_nct127	sasaeng udah pada gila emang dari dullu penjarain aja	0
2023-08-10T10:58:56Z	@zhangluobo2892	pulang kerja dengerin jaehyun langsung ilang capeknya	1
2023-08-08T09:18:03Z	@finachyni3834	jae kamu keren banget	1
2023-10-08T12:44:37Z	@user_w1j4	sm gaa adil banget masa jaehyun ga dikasih tour konser sendiri	0

memahami sentimen komentar pada video musik "Horizon" oleh Jung Jaehyun, yang akan dijabarkan lebih rinci pada tabel 6 confusion matrix dibawah ini.

Tabel 6. Confusion matrix

	Prediksi Positif (1)	Prediksi Negatif (0)
Actual Positif	460	0
Actual Negatif	14	0

Pada tabel 6 Confusion Matrix diatas dijelaskan bahwa actual positif dari 465 sampel positif sebenarnya, semuanya diprediksi positif. Sedangkan, actual negatif dari 14 sampel negatif sebenarnya, tidak ada yang diprediksi negatif, sedangkan 14 sampel tersebut diprediksi positif.



Gambar 4. Confusion Matrix Naïve Bayes

Dapat dijelaskan pada gambar 4 diatas, bahwa true positive (TP) : 465 (baris Asli 1, dikolom prediksi 1) dimana model memprediksi positif (1) dan sebenarnya juga positif (1), lalu False Negative (FN) : 0 (baris Asli, dikolom prediksi 0) dimana model memprediksi negatif (0) padahal sebenarnya positif (1). Selanjutnya adalah False Positive (FP) : 14 (baris Asli 0, dikolom prediksi 1) dimana model memprediksi positif (1) padahal sebenarnya negatif (0), lalu True Negative (TN) : 0 (baris Asli 0, dikolom prediksi 0) dimana model memprediksi positif (1) padahal sebenarnya negatif (0). Dapat di simpulkan jika semua prediksi model diarahkan ke kelas positif dengan menghasilkan False Positive sebanyak 14 kasus dan True Positive sebanyak 465 kasus.

Tabel 7. Klasifikasi Naïve Bayes

Metrik	Klasifikasi Naïve Bayes
Akurasi	0.97%
Presisi	0.00%
Recall	0.00%
F1-Score	0.00%

Pada tabel 7 diatas, dapat dijelaskan jika dari analisis menggunakan algoritma Naïve Bayes menghasilkan tingkat akurasi sebanyak 0.97% dan Presisi, Recall serta F1-Score menunjukkan

hasil 0.00% dimana nilai tersebut menunjukkan bahwa tidak ada prediksi positif yang benar. Dalam evaluasi confusion matrix, meski model memprediksi banyak data sebagai positif, semuanya dianggap salah ketika mempertimbangkan kelas negatif (karena tidak ada True Negative).

Perbandingan Algoritma

Dari hasil algoritma Naïve Bayes yang akurasi masih belum memuaskan, peneliti melakukan perbandingan metode kembali dengan membandingkan kinerja antara algoritma Naïve Bayes dan Support Vector Machine (SVM). Evaluasi dilakukan dengan menggunakan metrik kinerja utama, seperti akurasi, presisi, recall, F1-score, dan waktu komputasi. Pada tabel 8 dibawah ini akan dijabarkan hasil perbandingannya.

Tabel 8. Hasil Perbandingan Naïve Bayes dan SVM(Linier Kernel)

Metrik	Naïve Bayes	SVM (Linier Kernel)
Akurasi	0.97%	0.98%
Presisi Negatif (0)	0.00%	100%
Presisi Positif (1)	0.00%	0.98%
Recall Negatif (0)	0.00%	0.36%
Recall Positif (1)	0.00%	100%
F1-Score Negatif (0)	0.00%	0.53%
F1-Score Positif (1)	0.00%	0.99%

Pada tabel 8 diatas, Support Vector Machine (SVM) menunjukkan hasil akurasi yang jauh lebih tinggi dengan jumlah akurasi 0.98% dibandingkan dengan Naïve Bayes yang hanya menghasilkan akurasi berjumlah 0.97%.

Tabel 9. Hasil Uji Kernel SVM

Kernel	Prediksi	Accuracy	Precision	Recall	F1-Score
Linier	Negatif	0.98%	100%	0.36%	0.53%
	Positif		0.98%	100%	0.99%
Polynomial	Negatif	0.97%	0.00%	0.00%	0.00%
	Positif		0.97%	100%	0.98%
RBF	Negatif	0.97%	100%	0.09%	0.17%
	Positif		0.97%	100%	0.99%
Sigmoid	Negatif	0.98%	100%	0.36%	0.53%
	Positif		0.98%	100%	0.99%
Laplacian	Negatif	0.97%	0.00%	0.00%	0.00%
	Positif		0.97%	100%	0.98%
Exponential	Negatif	0.97%	0.00%	0.00%	0.00%
	Positif		0.97%	100%	0.98%
Multiquadric Inverse	Negatif	0.95%	0.00%	0.00%	0.00%
	Positif		0.97%	0.98%	0.97%
Multiquadric	Negatif	0.97%	0.00%	0.00%	0.00%
	Positif		0.97%	100%	0.98%
Wavelet	Negatif	0.97%	100%	0.09%	0.17%
	Positif		0.97%	100%	0.99%
Anova	Negatif	0.97%	0.00%	0.00%	0.00%
	Positif		0.97%	100%	0.98%

Pada tabel 9 diatas merupakan hasil perbandingan kernel SVM diatas, menunjukkan bahwa kernel linier dan sigmoid menghasilkan akurasi 0.98%, pada kernel linier precision kelas negatif tinggi (100%) yang berarti semua prediksi negatif berhasil terdeteksi, recall untuk kelas negatif rendah (0.36%) yang berarti banyak prediksi negatif salah (False Negative), F1-Score untuk kelas negatif tinggi (0.99%) yang berarti menunjukkan keseimbangan antara precision dan recall. Dan untuk kelas positif, model menghasilkan precision tinggi (0.98%) yang berarti sebagian dari data positif berhasil terdeteksi dengan baik, untuk recall kelas positif tinggi (100%) yang berarti semua data positif berhasil terdeteksi dan F1-Score untuk kelas positif (0.99%) yang mana menunjukkan keseimbangan hasil antara precision dan recall.

Hasil evaluasi algoritma Naïve Bayes menunjukkan akurasi sebesar 97%, dengan kelemahan utama pada prediksi kelas negatif, seperti terlihat pada confusion matrix. Tidak ada sampel aktual negatif yang terprediksi benar sebagai negatif. Sebaliknya, algoritma SVM dengan kernel linier memberikan akurasi tertinggi sebesar 98%, dengan kinerja yang lebih baik pada kelas negatif, meskipun recall untuk kelas negatif tetap rendah pada 36%. Hasil perbandingan kedua algoritma menunjukkan bahwa SVM unggul dalam menangani data tidak seimbang, sementara Naïve Bayes cocok untuk eksplorasi awal pada dataset kecil karena kesederhanaannya.

4. Pembahasan

Aplikasi Temuan untuk Strategi Pemasaran dan Pengelolaan Penggemar

Hasil penelitian ini memberikan wawasan yang dapat dimanfaatkan untuk memperkuat strategi pemasaran dan pengelolaan penggemar di media sosial. Algoritma SVM, dengan tingkat akurasi yang tinggi, memungkinkan analisis sentimen yang lebih andal. Informasi ini dapat digunakan untuk memahami respons penggemar terhadap konten tertentu, seperti video musik atau kampanye promosi. Komentar dengan sentimen positif dapat dimanfaatkan sebagai bahan untuk memperluas daya tarik pemasaran, sedangkan komentar dengan sentimen negatif dapat menjadi panduan untuk mengidentifikasi area yang memerlukan perbaikan.

Dalam pengelolaan penggemar, hasil ini dapat membantu meningkatkan keterlibatan penggemar melalui intervensi berbasis data. Komentar negatif, misalnya, dapat dijadikan dasar untuk memperbaiki pengalaman penggemar, seperti memberikan tanggapan terhadap keluhan yang sering muncul atau memperbaiki aspek yang kurang disukai. Dengan cara ini, perusahaan dapat menjaga hubungan yang lebih erat dengan komunitas penggemar.

Analisis Pola Kesalahan pada Kelas Negatif

Kelemahan utama yang ditemukan dalam penelitian ini adalah rendahnya kinerja algoritma dalam mengklasifikasikan sentimen negatif. Pada algoritma Naïve Bayes, seluruh komentar negatif diklasifikasikan sebagai positif, menunjukkan bahwa algoritma ini kurang mampu mengenali pola yang mencirikan sentimen negatif. Hal ini kemungkinan besar dipengaruhi oleh ketidakseimbangan data, di mana komentar positif mendominasi dataset. Sementara itu, algoritma SVM menunjukkan presisi tinggi untuk kelas negatif, tetapi recall-nya rendah (36%). Artinya, sebagian besar komentar negatif tidak terdeteksi oleh model. Masalah ini menyoroti perlunya penyesuaian pada metode pembelajaran, seperti penggunaan teknik penyeimbangan data, misalnya dengan *oversampling* pada kelas minoritas. Selain itu, analisis lebih lanjut terhadap fitur-fitur teks yang memengaruhi klasifikasi dapat membantu meningkatkan kinerja model dalam mengenali sentimen negatif. Penelitian ini menekankan pentingnya memperbaiki kinerja algoritma, terutama untuk kelas yang kurang terwakili, agar hasil analisis sentimen dapat digunakan lebih efektif untuk pengambilan keputusan strategis.

5. Penutup

Penelitian ini berhasil menganalisis sentimen komentar pada video musik K-pop "Horizon" oleh Jung Jaehyun mengaplikasikan algoritma Naïve Bayes dan membandingkannya dengan algoritma Support Vector Machine (SVM). Dari hasil evaluasi, SVM menghasilkan performa yang jauh lebih baik dibandingkan Naïve Bayes dalam hal akurasi, presisi, recall, dan F1-score, khususnya dalam mengelola data yang tidak seimbang. Meskipun hasil evaluasi menunjukkan performa SVM lebih unggul dalam hal akurasi, presisi, recall dan F1-Score, algoritma Naïve Bayes tetap mempunyai keunggulan tersendiri dalam kesederhanaannya dan kemampuannya untuk bekerja secara efisien pada dataset berukuran kecil. Keunggulan ini menjadikan Naive bayes sebagai pilihan ideal untuk analisis awal atau dalam situasi di mana sumber daya komputasi terbatas. Dengan pendekatan berbasis probabilistik, algoritma ini juga mudah diimplementasikan dan diinterpretasikan, sehingga cocok untuk aplikasi dengan data yang tidak terlalu kompleks atau memiliki distribusi sederhana.

Penelitian ini memberikan kontribusi signifikan bagi pihak-pihak yang terkait. **Industri musik** dapat memanfaatkan hasil analisis sentimen untuk memahami opini penggemar, meningkatkan strategi promosi, dan memperbaiki aspek yang kurang disukai. **Tim pemasaran** dapat menggunakan data ini untuk mengidentifikasi tren respons audiens dan menyusun kampanye yang lebih relevan. Selain itu, penelitian ini menjadi referensi bagi **akademisi dan peneliti** dalam mengembangkan metode analisis sentimen, khususnya dalam mengatasi tantangan data tidak seimbang.

Pada penelitian selanjutnya, diharapkan agar memperluas cakupan data untuk meningkatkan generalisasi hasil, menggunakan teknik penyeimbangan data seperti *SMOTE*, dan mengeksplorasi algoritma lain, seperti Random Forest atau *deep learning*. Selain itu, pihak terkait seperti agensi musik dan platform media sosial dapat memanfaatkan model ini untuk memantau dan memahami opini publik terhadap konten secara otomatis.

6. Referensi

- [1] K. Horasman, A. Rara Fauziyah, and C. Siahaan, "Analisis Penggunaan Media Online Kvibes di Kalangan Penggemar Korean Pop (K-Popers)," *Jurnal Ilmiah Multidisiplin*, vol. 3, no. 6, 2024.
- [2] H. Al Rasyid Harpizon *et al.*, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.
- [3] M. R. Yoanita, H. Setiawan, P. Lucky, and T. Irawan, "Analisis Fitur-Fitur Yang Mempengaruhi Jumlah Subscribers Youtube Menggunakan Algoritma Naive Bayes Classifier".
- [4] M. Persada Pulungan, A. Purnomo, A. Kurniasih, S. Tinggi Ilmu Manajemen dan Ilmu Komputer ESQ, and P. Korespondensi, "PENERAPAN SMOTE UNTUK MENGATASI IMBALANCE CLASS DALAM KLASIFIKASI KEPERIBADIAN MBTI MENGGUNAKAN NAIVE BAYES CLASSIFIER APPLICATION OF SMOTE TO OVERCOME CLASS IMBALANCE IN THE MBTI PERSONALITY CLASSIFICATION USING THE NAÏVE BAYES CLASSIFIER", doi: 10.25126/jtiik.2024117989.
- [5] G. K. Pati and E. Umar, "Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K-Nearest Neighbor," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2309, Oct. 2022, doi: 10.30865/mib.v6i4.4635.

- [6] M. F. Ansyah, Abd. Ghofur, and L. F. Lidimillah, "Analisis Sentimen Komentar Youtube Terhadap Tayangan #Terbaru! Temuan dan Masalah Ahlak Di Ponpes Al-zaytun Menggunakan Metode Naïve Bayes," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 2, pp. 847–856, Apr. 2024, doi: 10.33379/gtech.v8i2.4034.
- [7] A. R. Abdillah and F. N. Hasan, "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier," *SMATIKA JURNAL*, vol. 13, no. 01, pp. 117–130, Jul. 2023, doi: 10.32664/smatika.v13i01.750.
- [8] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 542, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [9] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter".
- [10] R. A. Firsttama, A. A. Arifiyanti, and D. S. Y. Kartika, "Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 2, pp. 282–285, Apr. 2024, doi: 10.47233/jteksis.v6i2.1263.
- [11] F. Syofiani, S. Alam, and M. I. Sulisty S, "Analisis Sentimen Penilaian Masyarakat Terhadap Childfree Berdasarkan Komentar di Youtube Menggunakan Algoritma Naïve Bayes," *Jurnal Teknologi Informatika dan Komputer*, vol. 9, no. 2, pp. 688–703, Sep. 2023, doi: 10.37012/jtik.v9i2.1661.
- [12] R. A. Raharjo, I. Made, G. Sunarya, D. Gede, and H. Divayana, "Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin Covid-19 Di Twitter," vol. 15, no. 2, pp. 456–464, 2022, [Online]. Available: <http://journal.stekom.ac.id/index.php/elkom/page456>
- [13] M. A. Laksono, I. A. Kautsar, and H. Setiawan, "Implementasi Payment Gateway pada Platform Freelance Digital Menggunakan Rest API," *SMATIKA JURNAL*, vol. 14, no. 01, pp. 135–145, Jun. 2024, doi: 10.32664/smatika.v14i01.1227.
- [14] M. I. Santoso and A. R. Dzikrillah, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Pengguna Terhadap Kinerja Sistem Transportasi Umum Jakarta Menggunakan Algoritma Naive Bayes," *Media Online*, vol. 4, no. 6, pp. 3032–3043, 2024, doi: 10.30865/klik.v4i6.1936.
- [15] N. Muchammad Shiddieqy Hadna, P. Insap Santosa, and W. Wahyu Winarno, "STUDI LITERATUR TENTANG PERBANDINGAN METODE UNTUK PROSES ANALISIS SENTIMEN DI TWITTER," 2016.
- [16] M. A. Jibrán, A. Eviyanti, Y. Findawati, F. Sains, and D. Teknologi, "Deteksi Ujaran Kebencian Menggunakan Metode Support Vector Machine (SVM)."
- [17] M. O. Charisma, M. F. Hamzah, M. Erwin, I. Nurbaiti, and F. Kurniawan, "Klasifikasi Sentimen Terhadap Kebijakan PHK 55 Ribu Karyawan oleh BT Group menggunakan Algoritma Klasifikasi Naive Bayes," 2024. [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>
- [18] S. Mulyani and R. Novita, "IMPLEMENTATION OF THE NAIVE BAYES CLASSIFIER ALGORITHM FOR CLASSIFICATION OF COMMUNITY SENTIMENT ABOUT DEPRESSION ON

- YOUTUBE,” *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, pp. 1355–1361, Oct. 2022, doi: 10.20884/1.jutif.2022.3.5.374.
- [19] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes,” vol. 9, no. 2, pp. 785–795, 2022.
- [20] M. Jonathan and Y. Nataliani, “Analisis Sentimen Penilaian Masyarakat Indonesia terhadap GeNose pada Komentar Youtube Menggunakan Metode Naïve Bayes.” [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/decartesian>
- [21] D. Nasien *et al.*, “Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, Dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes,” 2024.