

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

Oleh:

Addriana Fatma Putri Indah Sari

Ade Eviyanti

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Januari, 2025



Pendahuluan

K-pop sebagai bagian dari Hallyu telah menjadi fenomena global yang menyebarkan budaya Korea ke seluruh dunia. Penggemar K-pop aktif di platform seperti YouTube, yang menyediakan data opini publik melalui fitur komentar. Namun, analisis sentimen komentar ini menghadapi tantangan ketidakseimbangan data.

Penelitian ini menggunakan algoritma Naïve Bayes untuk menganalisis sentimen komentar pada video musik "Horizon" oleh Jung Jaehyun, dengan kategori positif dan negatif. Selain itu, algoritma ini dibandingkan dengan Support Vector Machine (SVM) untuk mengevaluasi performa keduanya. Studi ini mengisi gap penelitian sebelumnya dengan memberikan analisis sentimen pada data komentar K-pop, yang memiliki karakteristik opini penggemar spesifik.

Pertanyaan Penelitian (Rumusan Masalah)

Dari latar belakang yang telah dikemukakan, rumusan masalah dalam penelitian ini meliputi pengelolaan data komentar YouTube yang tidak terstruktur dan tidak seimbang untuk menghasilkan analisis sentimen yang akurat, efektivitas algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam menganalisis sentimen komentar pada video musik Horizon oleh Jung Jaehyun, kelemahan utama dari algoritma Naïve Bayes dan SVM dalam mengklasifikasikan komentar berbahasa Indonesia, khususnya dalam mengenali sentimen negatif, serta penerapan kedua algoritma tersebut untuk membantu memahami opini penggemar K-pop melalui platform YouTube.

Metode

1.Studi Literatur: Meninjau jurnal dan referensi terkait analisis sentimen untuk memahami metode yang relevan.

2.Pengumpulan Data: Data komentar pada video musik "Horizon" oleh Jung Jaehyun dikumpulkan menggunakan YouTube API melalui Python di Google Colaboratory.

3.Crawling Data: Komentar dikumpulkan dari kanal YouTube SM Town pada 28 Mei 2024.

4.Pra-Pemrosesan: Data dibersihkan melalui beberapa tahap, yaitu:

- **Cleaning:** Menghilangkan tanda baca, emotikon, dan angka.
- **Tokenisasi:** Memisahkan kalimat menjadi kata-kata.
- **Normalisasi:** Mengubah kata ke bentuk dasar.
- **Stopword Removal:** Menghapus kata-kata umum yang tidak relevan.

Metode

5. Pelabelan Data: Komentar dikategorikan secara manual sebagai positif atau negatif, dengan validasi dari ahli bahasa.

6. Klasifikasi: Data dianalisis menggunakan algoritma Naïve Bayes. Penelitian ini juga membandingkan kinerja Naïve Bayes dan SVM untuk menentukan algoritma yang paling efektif dalam analisis sentimen, khususnya pada data komentar berbahasa Indonesia yang tidak seimbang.

7. Evaluasi dan Implementasi: Model dievaluasi berdasarkan metrik akurasi, presisi, recall, dan F1-score menggunakan confusion matrix.

Hasil

- Sampel Dataset**

Data yang diperoleh merupakan data komentar youtube dari Music Video (MV) Jung Jaehyun yang berjudul “HORIZON” dengan data mentah berjumlah 13.271 dan data komentar hanya diambil yang berbahasa Indonesia yang berjumlah 2.391 data.

publishedAt	authorDisplayName	textDisplay
2024-09-03T13:05:41Z	@haura6480	Seger bet liat ni Cowo
2024-08-30T10:43:45Z	@octopus8978	â€™™Aku masih 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	ì¸ì•.,ì.,œAku harap semakin banyak yang tau lagu iniê° ë™ìì•¼
2024-08-29T10:50:13Z	@naajae_14	âì,cinta banget sama Jaehyun âì,âì,•
2023-08-21T16:01:55Z	@iom_nct127	ž ‘ë..., ì~~fì™œ ì•ë...,ëž~ë “àìœ¼ë©Sasaeng udah pada gila emang dari dullu, penjarain aja

Hasil

- **Pengumpulan Data dan Crawling Data**

Pengumpulan data terlebih dahulu melakukan login ke (<https://console.developers.google.com/>) untuk mendapatkan Youtube API (API Key). Setelah mendapatkan API Key proses berikutnya adalah crawling data.

- **Proses Crawling Data**

```
[ ] # isikan dengan api key Anda
    api_key = 'AIzaSyACGJWMjY2XcFo0wzUelxsYoQkgP1ezlE8'

    # Enter video id
    # contoh url video = https://youtu.be/Y6_vy82Vo70?si=k2Gs8ZXBtozr18tr
    video_id = "Y6_vy82Vo70" #isikan dengan kode / ID video

    # Call function
    comments = video_comments(video_id)

    comments
```

- **Sampel Hasil Crawling Data**

```
['2023-12-24T18:26:45Z', '@yunoisjaehyun', 'i am still here, jaehyun', 4],
['2023-12-24T18:01:15Z', '@helwamahri', 'Kamu keren banget sayangkuuuu', 0],
['2023-12-23T21:15:57Z',
 '@daliblose9061',
 'My horizon Jaehyunnahhhh!!! Yes!! Whoo woo!! Love ❤️ you thx 😊 👯 🥰 🙏 🙌',
 1],
['2023-12-23T08:00:11Z', '@papakoji9239', 'jjh1 soon inshaallah', 1],
['2023-12-22T13:21:04Z', '@abihoer-bn9dn', 'Suaranya selalu bikin candu', 0],
['2023-12-22T04:16:54Z', '@mybeelov3', 'Lagu penyemangat', 0],
['2023-12-21T09:59:04Z', '@loeyemma7773', '❤️❤️❤️❤️❤️', 0],
['2023-12-20T03:47:05Z', '@anxiousme.', 'Im hereeee', 0],
['2023-12-17T16:45:06Z', '@yasmimp4', 'JJH1 is coming ❤️', 1],
['2023-12-17T11:22:20Z', '@igigiuv', '재현이천재사랑해'. 0].
```

Hasil

- Pra-pemrosesan data meliputi cleaning untuk menghapus tanda baca, emotikon, dan angka; tokenisasi untuk memisahkan kalimat menjadi kata; normalisasi untuk mengubah kata ke bentuk dasar; serta stopwords removal untuk menghapus kata-kata umum yang tidak relevan. Dibawah ini hasil dari pra-pemrosesan data.

publishedAt	authorDisplayName	textDisplay	stem
2024-09-03T13:05:41Z	@haura6480	Seger bet liat ni Cowo	seger bet liat ni cowo
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	ìç(ì•.,ì.,œAku harap semakin banyak yang tau lagu iniê° ë™ì•¼	aku harap semakin banyak yang tau lagu ini
2024-08-29T10:50:13Z	@naajae_14	âai,cinta banget sama Jaehyun âai,âai,•	cinta banget sama jaehyun

Hasil

- **Pelabelan Data**

Komentar dikategorikan secara manual sebagai positif bernilai (1) atau negative bernilai (0), dengan validasi dari ahli bahasa Bernama Joko Susilo, S.Hum., M.Hum.

publishedAt	authorDisplayName	stem	Label
2024-09-03T13:05:41Z	@haura6480	seger bet liat ni cowo	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

Hasil

- **Klasifikasi**

Proses klasifikasi menggunakan algoritma Naïve Bayes dimulai dengan pelabelan data, yang kemudian diproses untuk memprediksi pola dan mengklasifikasikannya menjadi positif atau negatif. Metode ini sering digunakan dalam analisis sentimen karena hanya memerlukan data yang relatif kecil untuk estimasi parameternya. Selain itu, algoritma Naïve Bayes akan melalui tahap perbandingan dengan algoritma Support Vector machine (SVM).

```
[ ] # Membuat model Naive Bayes  
model = MultinomialNB()  
model.fit(X_train_vect, y_train)
```

```
: from sklearn.svm import SVC
```



▼ MultinomialNB ⓘ ?

MultinomialNB()

Hasil

- **Evaluasi dan Implementasi**

Tahap ini melibatkan evaluasi hasil klasifikasi menggunakan confusion matrix untuk membandingkan performa algoritma Naïve Bayes dan Support Vector Machine (SVM). Implementasi dilakukan dengan pemrograman Python di Google Colaboratory, menggunakan metrik seperti akurasi, presisi, recall, dan F1-score untuk analisis performa kedua algoritma.

- **Rumus Evaluasi Confusion Matrix**

	Prediksi Positif	Prediksi Negatif
Positif	TP	FP
Negatif	FN	TN

- **Penjelasan Rumus Evaluasi Confusion Matrix**

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

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

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

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

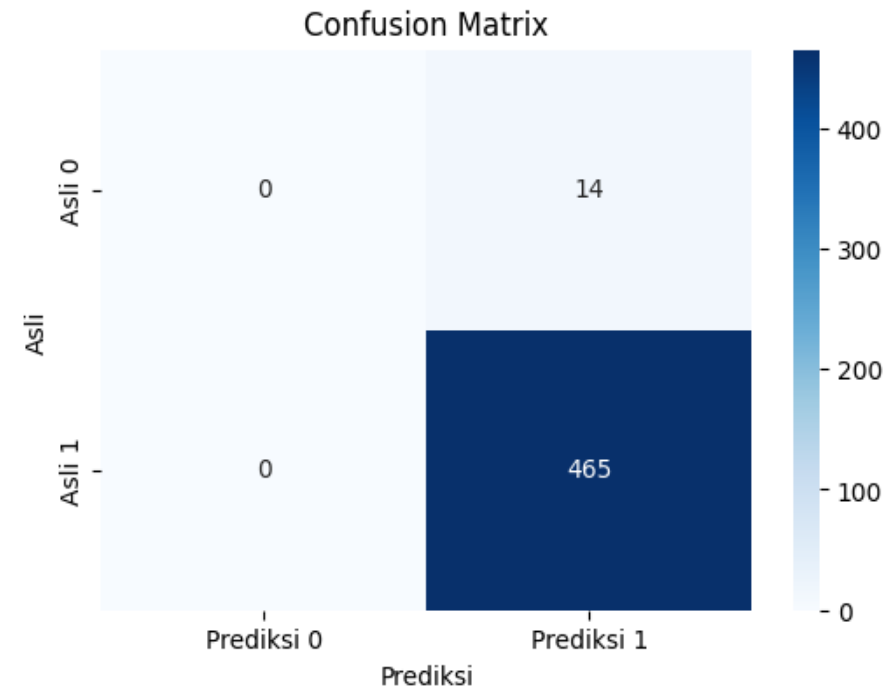
Pembahasan

- **Hasil Evaluasi Algoritma Naïve Bayes**

Analisis algoritma Naïve Bayes pada dataset komentar video musik "Horizon" oleh Jung Jaehyun menggunakan 80% data pelatihan dan 20% pengujian. Evaluasi kinerja berdasarkan akurasi, presisi, recall, dan F1-score menunjukkan efektivitas model. Confusion Matrix mengungkapkan 465 sampel positif diprediksi benar (TP: 465), 14 sampel negatif salah diprediksi sebagai positif (FP: 14), tanpa True Negative (TN: 0) atau False Negative (FN: 0). Model ini memprediksi semua data sebagai positif.

- **Tabel hasil dan Visualisasi Confusion Matrix**

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



Pembahasan

- **Lanjutan Hasil Algoritma Naïve Bayes**

Analisis dengan algoritma Naïve Bayes menghasilkan akurasi sebesar 0,97%. Namun, nilai presisi, recall, dan F1-score adalah 0,00%, menunjukkan tidak ada prediksi positif yang benar.

Dalam evaluasi confusion matrix, meskipun model memprediksi banyak data sebagai positif, semua prediksi dianggap salah untuk kelas negatif karena tidak ada True Negative.

- **Tabel hasil Akurasi Algoritma Naïve Bayes**

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

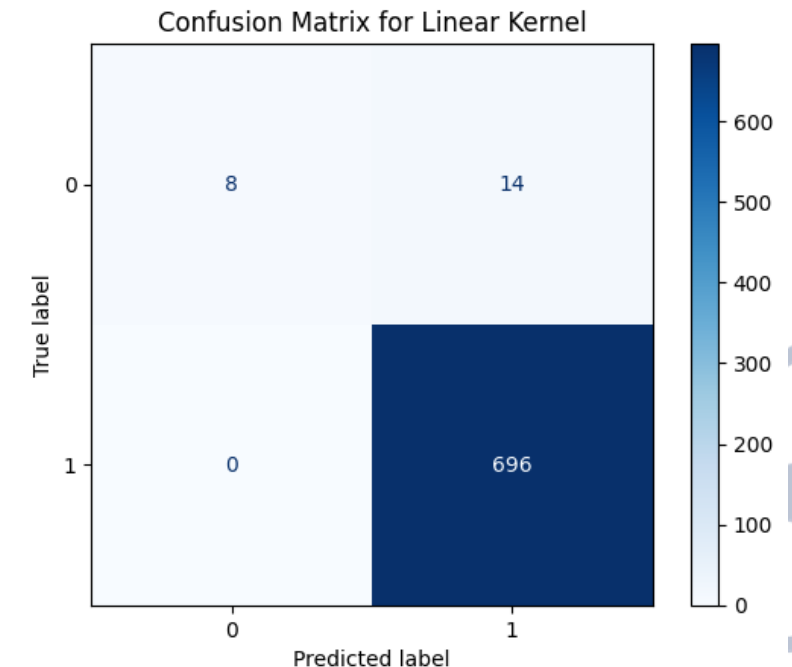
Pembahasan

- **Hasil Perbandingan Algoritma Naïve Bayes dan Support Vector Machine (SVM)**

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 confusion matrix seperti akurasi, presisi, recall, F1-score. Support Vector Machine (SVM) menunjukkan hasil akurasi yang jauh lebih tinggi dengan kernel Linier berjumlah akurasi 0.98% dibandingkan dengan Naïve Bayes yang hanya menghasilkan akurasi berjumlah 0.97%.

- **Tabel hasil Perbandingan Naïve Bayes dan SVM (Kernel Linier)**

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%



Manfaat Penelitian

Penelitian ini bermanfaat secara akademik dengan memperkaya studi analisis sentimen, khususnya pada data tidak seimbang berbahasa Indonesia, serta membandingkan performa algoritma Naïve Bayes dan SVM. Secara praktis, hasilnya dapat membantu industri musik memahami opini penggemar melalui komentar YouTube dan mendukung pengembangan alat pemantauan sentimen di media sosial. Selain itu, penelitian ini berkontribusi pada pengembangan teknologi analisis data dan mendukung pengambilan keputusan berbasis data dalam strategi pemasaran dan interaksi digital.

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 KEPRIBADIAN MBTI MENGGUNAKAN NAIVE BAYES CLASSIFIER APPLICATION OF SMOTE TO OVERCOME CLASS IMBALANCE IN THE MBTI PERSONALITY CLASSIFICATION USING THE NAIVE 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/jfeksis.v6i2.1263.
- [11] F. Syofiani, S. Alam, and M. I. Sulistyo 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.

Referensi

- [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. Dzibrillah, "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 Shiddiegy 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.unsraf.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.

