JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i3.7362

ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP ANIME SPY X FAMILY BAHASA INDONESIA MENGGUNAKAN NAÏVE BAYES

Nurmala Arita[1*], Nono Heryana[2], Azhari Ali Ridha[3]

[1,2,3]Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361; (0267) 641177

Keywords:
Analisis Sentimen;
Naïve Bayes;
KDD;
Youtube;
Spy x Family

Corespondent Email:
2110631250016@student.unsika.ac.id

Abstrak. Penelitian ini bertujuan untuk menganalisis sentimen pada komentar pengguna YouTube terhadap anime Spy x Family Bahasa Indonesia di channel Muse Indonesia. Algoritma yang digunakan pada penelitian ini yaitu algoritma Naive Bayes yang berfungsi untuk mengoptimalkan klasifikasi sentimen berdasarkan komentar dari YouTube. Knowledge Discovery in Database (KDD) digunakan sebagai metodologi penelitian yang terdiri dari 6 proses yaitu Data Selection, Data Cleaning, Tokenizing, Normalization, Stemming, dan Stop-word Removal. Pada penelitian ini, algoritma Naive Bayes menghasilkan kinerja yang baik pada hasil evaluasinya. Hasil dari penelitian menerangkan bahwa algoritma Naive Bayes menunjukkan hasil yang cukup signifikan dengan sentimen 57,83% bersifat positif dan 42,17% bersifat negatif. Adapun hasil performa dari algoritma Naive Bayes dengan membandingkan tiga rasio (60:40, 70:30, 80:20) untuk data latih dan data uji menghasilkan akurasi yang tinggi. Rasio 80:20 menghasilkan nilai tertinggi dengan akurasi mencapai 86,03%, presisi 81,08%, recall 98,90%, dan F1-Score 89,11%. Hasil ini telah memberi gambaran bahwa sebagian besar pengguna YouTube yang menonton Spy x Family Bahasa Indonesia memberi respons positif terhadap kehadiran anime tersebut dengan versi dubbing Indonesia.

Abstract. *This research aims to analyze sentiment on YouTube user comments on the Indonesian Spy x Family anime on the Muse Indonesia channel. The algorithm used in this research is the Naive Bayes algorithm which functions to optimize sentiment classification based on comments from YouTube. Knowledge Discovery in Database (KDD) is used as a research methodology consisting of 6 processes, namely Data Selection, Data Cleaning, Tokenizing, Normalization, Stemming, and Stop-word Removal. In this research, Naive Bayes algorithm produces good performance in the evaluation results. The results of the study explained that the Naive Bayes algorithm showed significant results with 57.83% positive sentiment and 42.17% negative. The performance results of the Naive Bayes algorithm by comparing three ratios (60:40, 70:30, 80:20) for training data and test data resulted in high accuracy. The 80:20 ratio produces the highest value with accuracy reaching 86.03%, precision 81.08%, recall 98.90%, and F1-Score 89.11%. These results have illustrated that most YouTube users who watched Spy x Family Indonesian gave a positive response to the presence of the anime with the Indonesian dubbed version.*

1. PENDAHULUAN

YouTube menjadi media pilihan sebagian besar masyarakat Indonesia untuk menonton, mengunggah, dan berbagi video. Radio Republik Indonesia atau rri.co.id menerangkan bahwa pengguna YouTube di Indonesia mencapai 139 juta pengguna (mencapai 53,8% dari populasi keseluruhan). Fenomena anime telah meraih popularitas tinggi di kalangan masyarakat Indonesia, terutama para remaja [1]. Saat ini, anime telah tersedia di platform YouTube dengan beberapa pilihan subtitle bahasa, salah satunya anime Spy x Family Bahasa Indonesia yang tersedia di channel youtube Muse Indonesia. Komentar pengguna YouTube terhadap anime tersebut semakin beragam, namun belum dianalisis secara sistematis untuk mengetahui sentimennya apakah cenderung positif atau negatif. Dalam ilmu komputer, terdapat cabang ilmu yang mempelajari tentang data, yaitu Data Mining. Data Mining merupakan sebuah proses untuk menemukan suatu pola dalam kumpulan data yang sebelumnya tidak diketahui lewat teknik analisis data [2]. Penelitian ini menerapkan salah satu algoritma dari bidang Data Mining, yaitu algoritma Naive Bayes. Algoritma Naive Bayes sendiri merupakan sebuah algoritma untuk klasifikasi yang cukup banyak digunakan pada Data Mining [3]. Algoritma ini dipilih karena terdapat beberapa penelitian sebelumnya yang mendapatkan nilai akurasi tinggi menggunakan Naive Bayes. Salah satunya seperti pada penelitian [4], dari penelitian tersebut, didapatkan hasil akurasi Naïve Bayes sebesar 90%, Decision Tree sebesar 83%, dan Random Forest sebesar 87%. Yang mana Naive Bayes meraih nilai akurasi tertinggi. Oleh karena itu, efektivitas Naive Bayes dalam mengklasifikasikan sentimen dari pengguna YouTube yang menonton anime Spy x Family Bahasa Indonesia perlu dibuktikan. Signifikansi penelitian bertujuan mengisi kesenjangan pengetahuan tentang analisis sentimen dalam konteks hiburan dan memberi kontribusi nyata bagi para pengambil kebijakan tentang promosi budaya populer asing di Indonesia dalam memahami pandangan pengguna terhadap hiburan yang mereka sajikan.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen yaitu adalah sebuah metode yang digunakan untuk menyaring atau memisahkan data dari opini tertentu serta mengolah dan memahami teks dari data secara otomatis untuk melihat isi sentimen yang terkandung di dalam sebuah opini. Analisis sentimen bisa digunakan untuk membeberkan pemikiran ataupun opini masyarakat terhadap suatu persoalan [5]. Analisis sentimen memiliki kelebihan penting, yaitu dapat menghemat tenaga dan waktu saat melakukan penelitian dengan jumlah data yang terbilang besar. Penerapan analisis sentimen berbahasa Indonesia meliputi beberapa macam seperti layanan pelanggan, penilaian produk, serta pengamatan opini dari pengguna media sosial [6].

2.2. YouTube

YouTube merupakan media sharing (media untuk berbagi) dan salah satu jenis media sosial yang menunjang para penggunanya untuk berbagi video. Pengguna YouTube dibebaskan untuk berkomentar, memberi pendapat atau opini dari konten-konten yang disajikan. Seiring perkembangan YouTube hingga saat ini, konten yang disediakan pun juga semakin banyak dan bervariasi [7]. Untuk mengambil data komentar YouTube, digunakan YouTube API v3, sebuah tools yang memungkinkan penggunanya untuk mengakses data yang didapat dari YouTube secara programatik.

2.3. Anime

Anime adalah sebutan untuk animasi yang berasal dari negeri sakura, yaitu Jepang. Anime biasanya dibuat dari gambar tangan dan gambar digital yang dibuat menggunakan komputer. Orang yang membuat ataupun berkontribusi dalam pembuatan anime disebut animator. Media yang banyak dimanfaatkan oleh masyarakat Indonesia ketika menonton anime adalah situs film sebesar 64.8%, melalui YouTube sebesar 48.8%, dan melalui internet sebesar 24.4% [8].

2.4. Naïve Bayes

Naive Bayes merupakan algoritma klasifikasi yang cukup sederhana tetapi efisien. Naive Bayes biasa diterapkan untuk memberi prediksi pada suatu kasus seperti rekomendasi produk dan digunakan juga pada bidang medis untuk mendiagnosis suatu penyakit [9]. Dengan

menggunakan fitur-fitur yang diperoleh dari teks, algoritma Naive Bayes ini dapat memberikan prediksi sentimen yang akurat [10].

2.5. Data Mining

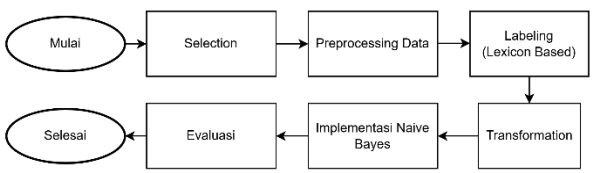
Data mining merupakan sebuah proses untuk menambang informasi penting dari data. Disebut juga sebagai proses menemukan dan menyaring pengetahuan atau informasi yang diperlukan menggunakan algoritma tertentu sesuai dengan informasi yang dicari [11]. Data mining memiliki beberapa metode seperti Knowledge Discovery in Database (KDD), dan SEMMA (Sample, Explore, Modify, Model, Assess).

2.6. Knowledge Discovery Database (KDD)

Knowledge Discovery in Database merupakan salah satu proses pada Data Mining yang memiliki tujuan untuk mencari dan menghasilkan insight baru dari suatu data untuk landasan dalam pengambilan keputusan. KDD bekerja dengan cara menggali pola saat memperoleh informasi dari suatu data dengan cara menggunakan suatu algoritma. KDD juga memiliki alur atau beberapa tahapan, seperti Data Selection (menyeleksi data), Data Preprocessing (pra-pemrosesan data), Data Transformation (transformasi data), Data Mining (penambangan data), dan Evaluation [12].

3. METODE PENELITIAN

Metodologi yang diimplementasikan adalah metode Knowledge Discovery Databases (KDD) karena penelitian akan berfokus pada analisis data teks menggunakan teknik pemrosesan bahasa alami (NLP). Tahapan KDD yang digunakan terdiri dari Selection dan Preprocessing. Dilanjutkan dengan labeling (Lexicon-based), transformation, implementasi Naive Bayes lalu evaluasi menggunakan Confusion Matrix dan Wordcloud.



Gambar 1. Alur Penelitian

3.1. Selection

Selection dilakukan untuk pengumpulan data yang diawali dengan mengumpulkan komentar dari pengguna youtube yang terkait dengan anime Spy x Family Bahasa Indonesia. Data dikumpulkan dengan cara Crawling data komentar anime Spy x Family Bahasa Indonesia yang berasal dari episode satu sampai dua belas pada season pertama menggunakan YouTube Data API v3.

3.2. Preprocessing

Preprocessing Data dilakukan untuk meningkatkan kualitas data menggunakan metode KDD, mencakup proses cleaning untuk pembersihan data sekaligus mengubah huruf kapital menjadi huruf kecil, tokenizing untuk memecah teks komentar menjadi kata, normalization untuk mengubah kata tidak baku menjadi baku, stemming untuk mengurangi kata-kata menjadi bentuk dasar, stop-word removal untuk menghapus kata-kata umum yang tidak memiliki makna signifikan.

3.3. Labeling

Pada langkah ini yang dilakukan adalah memberi label pada dataset berupa label sentimen positif dan negatif. Proses labeling memanfaatkan Lexicon yang memuat kamus kata opini. Metode Lexicon Based merupakan salah satu cara untuk menganalisis sentimen di sosial media dengan memanfaatkan kamus sebagai sumber bahasa untuk mengklasifikasikan sentimen dari tiap opini sehingga kalimat sentimen yang ada dapat diklasifikasikan menjadi kelas positif ataupun negative [13]. Proses labeling dilakukan setelah preprocessing karena untuk membersihkan data sebelum data tersebut dibaca sistem untuk diberi label sesuai dengan jurnal [14].

3.4. Transformation

Tahap ini mencari frekuensi kemunculan setiap kata (term frequency) yang akan menghasilkan matriks dari TF-IDF dengan cara mengubah dataset yang awalnya berbentuk string (teks) menjadi numerik (angka) agar dapat digunakan sebagai input ke dalam algoritma Naïve Bayes di tahap selanjutnya. TF-IDF dipilih karena sangat populer dan dapat mempertimbangkan keunikan juga frekuensi kata [15]

3.5. Implementasi Naïve Bayes

Pada langkah ini yang dilakukan adalah menerapkan Naive Bayes lalu melatih model pada data penelitian. Setelah itu akan dilakukan uji model pada data pengujian untuk evaluasi kinerja.

3.6. Evaluasi

Evaluasi diawali dengan melakukan tiga skenario lalu membagi data menjadi data latih dan uji dengan rasio 60:40, 70:30, dan 80:20 menggunakan Confusion Matrix. Lalu mencari akurasi, presisi, recall, dan F1-Score. Setelah itu, akan disajikan wordcloud sebagai representasi dari kata apa saja yang dominan muncul di dalam dataset yang digunakan agar mendapat pemahaman dengan lebih mudah.

4. HASIL DAN PEMBAHASAN

4.1. Selection

Dilakukan pengumpulan komentar menggunakan teknik crawling pada YouTube API v3 melalui Python. Data yang digunakan adalah komentar yang diambil dari YouTube pada bulan Januari 2025, disaat anime tersebut sedang ramai diperbincangkan penontonnya. Data komentar kemudian disimpan kedalam format file csv.

authorDisplayName	textDisplay
@MuseIndonesia	Kalian suka Anya? 😘 Tonton semua episodenya di tautan
@Muhamadilham-uc1sp	anya lucu banget boleh aku bawa pulang ngga min😂
@Starboy-x9w	Suka lucu
@husni2329	Sama😂
@cokrimecokrime7529	Li pi
@ArjunaPutra-m1l	Suka lucu +gemoyyy
@MairaSiregar-p1x	■■✿✿✿✿■■ ■✿✿✿✿✿✿✿■ ✿
@MairaSiregar-p1x	{__/} {__/} (•.•) (•-•) I LOVE YOU / >💜 💜 < \
@CarrolaSaputri	Anya imut bangat 😊😊
@MojoMojo-k4h	Nona cilik nggak tuh
@videoandro1582	Suka🎉 😊
@Khoirul28-o6b	Tahun brp ni filem😅
@nusurfatika4486	Suka
@nusurfatika4486	Ngk
@adityakrishnawestjava9201	Saya akan boyong spy x family ke RTV ato tv indo lainnya!

Gambar 2. Hasil Crawling

4.2. Preprocessing

Tahap ini terdiri dari cleaning, tokenizing, normalization, stemming, dan stop-word removal. Berikut tahapannya:

4.2.1. Cleaning

Tahap ini melibatkan penghapusan tanda baca, angka, simbol-simbol yang kurang relevan, dan case folding (mengubah huruf besar ke kecil).

Tabel 1. Hasil Cleaning

Sebelum	Sesudah
Kalian suka Anya? 😘 Tonton semua episod...	kalian suka anya tonton semua episodenya di ta...
anya lucu banget boleh aku bawa pulang ngga min😂	anya lucu banget boleh aku bawa pulang ngga min
Sama😂	sama
Suka lucu +gemoyyy	suka lucu gemoyyy
Anya imut bangat 😊😊	anya imut bangat

4.2.2. Tokenizing

Tiap kalimat dalam komentar dipecah menjadi satu kata agar data dapat dianalisis perkata. Proses ini tentunya akan memudahkan tahapan yang akan datang.

Tabel 2. Hasil Tokenizing

Sebelum	Sesudah
kalian suka anya tonton semua episodenya di ta...	[kalian, suka, anya, tonton, semua, episodenya...
anya lucu banget boleh aku bawa pulang ngga min	[anya, lucu, banget, boleh, aku, bawa, pulang,...
suka lucu	[suka, lucu]
suka lucu gemoyyy	[suka, lucu, gemoyyy]
anya imut bangat	[anya, imut, bangat]

4.2.3. Normalization

Tahap ini bertujuan untuk mengubah kata-kata yang bentuknya tidak baku menjadi baku menggunakan kamus normalisasi.

Tabel 3. Hasil Normalization

Sebelum	Sesudah
[anya, lucu, banget, boleh, aku, bawa, pulang,...	[anya, lucu, sekali, boleh, aku, bawa, pulang,...
[suka, lucu, gemoyyy]	[suka, lucu, gemas]
[anya, imut, bangat]	[anya, imut, sekali]
[nona, cilik, nggak, tuh]	[nona, cilik, tidak, itu]

4.2.4. *Stemming*

Tahap yang mengubah kata-kata menjadi bentuk dasar dengan menghilangkan imbuhan seperti awalan, akhiran, beserta gabungannya. Proses stemming yang dilakukan menggunakan bahasa pemrograman python dengan bantuan library Sastrawi.

Tabel 4. Hasil Stemming

Sebelum	Sesudah
yang difoto botak sebelah astaga	yang foto botak belah astaga
kalau tidak salah ini anime latar belakangnya di jerman	kalau tidak salah ini anime latar belakang jerman
aku suka ketika aku mendengar di wasaap	aku suka ketika aku dengar di wasaap
aku b aja ketika lihat filmnya	aku b aja ketika lihat film
pengisi suara anya nya cocok sangat	isi suara anya nya cocok sangat

4.2.5. *Stop-word Removal*

Tahap ini menghilangkan kata-kata yang tidak memiliki makna signifikan untuk menyederhanakan dan mengurangi noise yang ada sehingga hanya kata relevan yang benar-benar mewakili opini atau sentimen pengguna yang dicari.

Tabel 5. Hasil Stop-word

Sebelum	Sesudah
anya lucu sekali boleh aku bawa pulang tidak min	anya lucu bawa pulang min
anya imut sekali	anya imut
nona cilik tidak itu	nona cilik
tahun berapa ni film	ni film
yang foto botak belah astaga	foto botak belah astaga

4.3. *Labeling*

Tahap ini menggunakan Lexicon kamus sentimen positif negatif yang berasal dari kamus kata opini yang dibuat oleh Liu dan telah diterjemahkan oleh masdevid di Github, kemudian diinput pada python.

textDisplay	sentiment
<a href="https://www.youtube.com/watch?v=kCjxV...	negative
sad juga bapak anya apakah bapak anya yatim<br...	negative
Singkat saja "ayah bunda mau sun sun an&q...	negative
Kalau di tv tayang di MNCTV wih keren dah 🎉	positive
Mukanya damyan memerah itu tandanya dia suka 😊 ❤...	positive
Loid paling keren di keluarga besar 🎊 🎊 🎊 🎊	positive
Keren banget 🎉	positive
Bjir senyuman anya 😅 😅	positive
Cerewet banget demian sampai bilang cereweeeee...	negative
Ya ampun yuri yord 😅 😂	negative

Gambar 3. Hasil Labeling

4.4. *Transformation*

Tahap ini mengubah teks (huruf) menjadi bentuk numerik (angka). Transformation menggunakan perhitungan seberapa sering suatu kata muncul serta memboboti tiap kata berdasarkan keunikannya pada seluruh komentar.

yuriiya	yuripermisi	yuristasiun	yuritoit	yutup	zina	zom	zombie	zzz	zzzz
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 4. Hasil TF-IDF

	top_term	score
0	bahasa	0.581815
1	bangat	0.826206
2	difoto	0.548502
3	aku	0.571135
4	tembok	0.440726

Gambar 5. Kata Nilai TF-IDF Tertinggi

4.5. *Implementasi Naïve Bayes*

Implementasi Naïve Bayes dilakukan dengan mengimport library yang dibutuhkan seperti train test split dan multinomialNB.

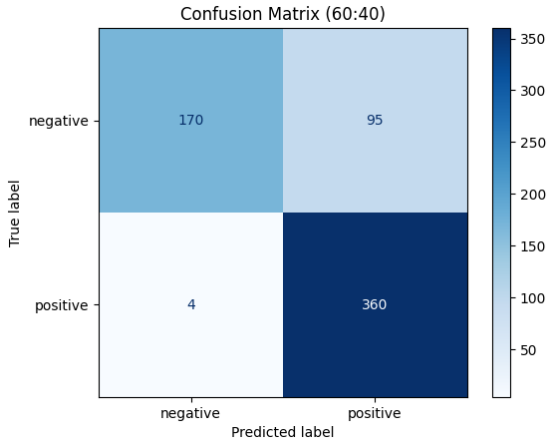
```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report
```

Gambar 6. Implementasi Library Naïve Bayes

4.6. Evaluasi

Dibuat Confusion Matrix untuk pengujian model Naive Bayes dengan tiga skenario pembagian data latih dan data uji, yaitu 60:40, 70:30, dan 80:20. Setelah itu, dibuatlah wordcloud.

4.6.1. Confusion Matrix 60:40



Gambar 7. Confusion Matrix 60:40

True Positive: 360	False Positive: 95
True Negative: 170	False Negative: 4

a. Akurasi

$$\frac{TP+TN}{TP+TN+FP+FN} = \frac{360+170}{360+170+95+4} = 0.8426 \quad (1)$$

b. Presisi

$$\frac{TP}{TP+FP} = \frac{360}{360+95} = 0.7912 \quad (2)$$

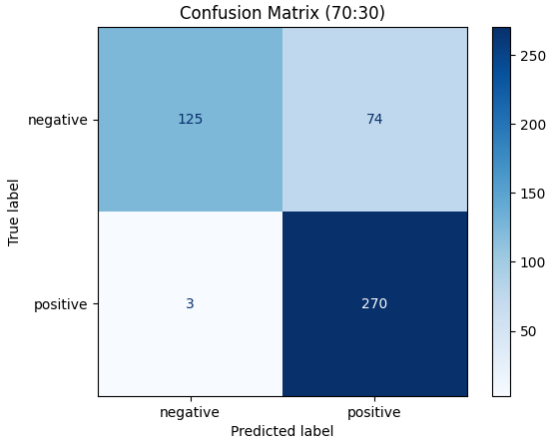
c. Recall

$$\frac{TP}{TP+FN} = \frac{360}{360+4} = 0.9890 \quad (3)$$

d. F1-Score

$$2 \times \frac{Presisi \times Recall}{Presisi+Recall} = 2 \times \frac{0.7912 \times 0.9890}{0.7912+0.9890} = 0.8791 \quad (4)$$

4.6.2. Confusion Matrix 70:30



Gambar 8. Confusion Matrix 70:30

True Positive: 270	False Positive: 74
True Negative: 125	False Negative: 3

a. Akurasi

$$\frac{TP+TN}{TP+TN+FP+FN} = \frac{270+125}{270+125+74+3} = 0.8369 \quad (1)$$

b. Presisi

$$\frac{TP}{TP+FP} = \frac{270}{270+74} = 0.7849 \quad (2)$$

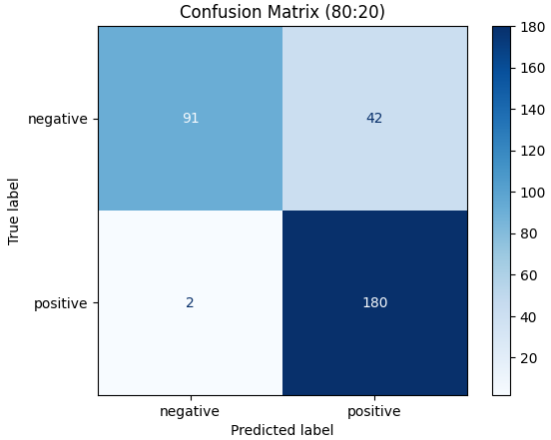
c. Recall

$$\frac{TP}{TP+FN} = \frac{270}{270+3} = 0.9890 \quad (3)$$

d. F1-Score

$$2 \times \frac{Presisi \times Recall}{Presisi+Recall} = 2 \times \frac{0.7849 \times 0.9890}{0.7849+0.9890} = 0.8752 \quad (4)$$

4.6.3. Confusion Matrix 80:20



Gambar 9. Confusion Matrix 80:20

True Positive: 180	False Positive: 42
True Negative: 91	False Negative: 2

a. Akurasi

$$\frac{TP+TN}{TP+TN+FP+FN} = \frac{180+91}{180+91+42+2} = 0.8603 \quad (1)$$

b. Presisi

$$\frac{TP}{TP+FP} = \frac{180}{180+42} = 0.8108 \quad (2)$$

c. Recall

$$\frac{TP}{TP+FN} = \frac{180}{180+2} = 0.9890 \quad (3)$$

d. F1-Score

$$2 \times \frac{Presisi \times Recall}{Presisi+Recall} = 2 \times \frac{0.8108 \times 0.9890}{0.8108+0.9890} = 0.8911 \quad (4)$$

4.6.4. Wordcloud

Gambar 10. Wordcloud

Berdasarkan visualisasi ini, tampak kata "anya", "yor", "loid, "lucu", "suka", dan "anime" merupakan kata yang paling dominan. Hal ini mengindikasikan bahwa karakter utama (Anya, Yor, dan Loid) serta kesan positif seperti "lucu", "suka", "seru", "tertawa" mendominasi persepsi pengguna. Visualisasi tersebut memberi gambaran penting bahwa mayoritas komentar bersifat positif dan berfokus pada tokoh serta kesan menyenangkan dari Spy x Family Bahasa Indonesia yang mereka tonton.

5. KESIMPULAN

a. Berdasarkan hasil penelitian, penerapan algoritma Naive Bayes untuk klasifikasi sentimen komentar YouTube terhadap Anime Spy x Family Bahasa Indonesia menghasilkan sentimen yang cukup signifikan dengan data 57,83% bersifat positif dan 42,17% bersifat negatif. Adapun hasil performa dari algoritma Naive Bayes dengan membandingkan tiga rasio (60:40, 70:30, 80:20) menghasilkan akurasi yang tinggi. Rasio 80:20 menghasilkan nilai tertinggi yaitu akurasi mencapai 86,03%, presisi 81,08%, recall 98,90%, dan F1-Score 89,11%. Hasil ini memberikan gambaran umum bahwa sebagian besar pengguna memberikan respons positif terhadap kehadiran anime Spy x Family versi dubbing Indonesia.

b. Komentar positif cenderung berisi banyak pujian penonton yang sangat menikmati alur cerita Spy x Family dengan dubbing Bahasa Indonesia dan juga pujian untuk pengisi suara karakter utama pada anime tersebut, yaitu Anya Forger, Loid Forger, dan Yor Forger. Sedangkan komentar negatif cenderung berisi ujaran kata atau kalimat yang kurang baik terhadap karakter yang ada.

6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

7. DAFTAR PUSTAKA

[1] M. Rizal *et al.*, "Analisis Sentimen Pengguna Twitter (X) Terkait Film One Piece Menggunakan Metode Algoritma Naive Bayes," *Jurnal Sistem Informasi Kaputama (JSIK)*, vol. 8, no. 1, 2024.

[2] A. De Wibowo Muhammad Sidik, I. Himawan Kusumah, A. Suryana, M. Artiyasa, and A. Pradiftha Junfithrana, "Gambaran Umum Metode Klasifikasi Data Mining," vol. 2, no. 2, pp. 34–38, 2020.

[3] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1.

[4] M. Yasir and R. Suraji2, "Perbandingan Metode Klasifikasi Naïve Bayes, Decision Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube."

[5] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter", *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available:

http://jim.teknokrat.ac.id/index.php/informatika

[6] M. Amien, "Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2304.02746

[7] D. A. Rahman, R. B. Waskitho, M. Fajrul, A. U. Nuha, and N. A. Rakhmawati, "Klasterisasi Topik Konten Channel Youtube Gaming Indonesia Menggunakan Latent Dirichlet Allocation."

[8] Y. Toi, "Kepopuleran dan Penerimaan Anime Jepang Di Indonesia," *Ayumi : Jurnal Budaya, Bahasa dan Sastra*, vol. 7, no. 1, Jul. 2020, doi: 10.25139/ayumi.v7i1.2808.

[9] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," *Soft comput*, vol. 25, no. 3, pp. 2277–2293, Feb. 2021, doi: 10.1007/s00500-020-05297-6.

[10] A. Setiawan and R. R. Suryono, "Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 183–192, Jun. 2024, doi: 10.29408/edumatic.v8i1.25667.

[11] F. Handayani, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," *Jurnal Teknologi dan Informasi*, doi: 10.34010/jati.v12i1.

[12] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4902.

[13] M. Bagas, D. Putra, and E. Setiawan, "Metode Lexicon Based Untuk Analisis Sentimen Pengguna Twitter Terhadap Kinerja Isp (Studi Kasus : Indihome, Biznet, Myrepublic)," 2024. [Online]. Available: http://dev.twitter.com

[14] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," 2022. [Online]. Available: https://github.com/evanmartua34/

[15] I. S. Wibowo, A. Witanti, and I. Susilawati, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF", [Online]. Available: http://jurnal.mdp.ac.id