

ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA SITUS IMDB MENGUNAKAN METODE NAIVE BAYES

FERY ANUAR RAMADHAN PUTRA¹, FRIDO FIRMAN FADILAH², ULTACH ENRI³
Universitas Singaperbangsa Karawang^{1, 2, 3}
feryanuar24@gmail.com¹

ABSTRACT

This research aims to analyze the accuracy of Oppenheimer film sentiment based on audience reviews written via the Internet Movie Database (IMDb) website using the Naive Bayes method. Audience reviews on the IMDb site are a valuable source of information in understanding audience opinions and responses to a film. In this research, researchers implemented the Naive Bayes algorithm classification method to classify reviews as positive or negative sentiment. Movie review data from IMDb is collected and entered into the pre-processing stage, then relevant features are extracted to train the Naive Bayes model. The evaluation results show that the Naive Bayes method can recognize sentiment in Oppenheimer film reviews with a significant level of accuracy. The findings of this research provide valuable insight for the film industry in understanding audience responses to these films, and the sentiment information obtained can be used as a basis for better decision making in film development and marketing. However, researchers acknowledge that there are limitations, especially in classification accuracy in reviews that use ambiguous or unclear language. Therefore, future research could involve other methods or combine several methods to improve the accuracy and reliability of sentiment analysis of film reviews.

Key Words: IMDb, Movie Reviews, Naive Bayes, Sentiment Analysis

ABSTRAK

Penelitian ini bertujuan untuk menganalisis keakuratan sentimen film Oppenheimer berdasarkan ulasan penonton yang ditulis melalui situs web Internet Movie Database (IMDb) menggunakan metode Naive Bayes. Ulasan penonton di situs IMDb merupakan sumber informasi yang berharga dalam memahami pendapat dan tanggapan penonton terhadap suatu film. Dalam penelitian ini, peneliti mengimplementasikan metode klasifikasi algoritma Naive Bayes untuk mengklasifikasikan ulasan sebagai sentimen positif atau negatif. Data ulasan film dari IMDb dikumpulkan dan masuk ke tahap pre-processing, kemudian fitur-fitur yang relevan diekstraksi untuk melatih model Naive Bayes. Hasil evaluasi menunjukkan bahwa metode Naive Bayes dapat mengenali sentimen dalam ulasan film Oppenheimer dengan tingkat akurasi yang signifikan. Temuan penelitian ini memberikan wawasan berharga bagi industri film dalam memahami respons penonton terhadap film ini, dan informasi sentimen yang diperoleh dapat digunakan sebagai dasar untuk pengambilan keputusan yang lebih baik dalam pengembangan film dan pemasaran. Meskipun demikian, peneliti mengakui adanya keterbatasan, terutama dalam akurasi klasifikasi pada ulasan yang menggunakan bahasa yang ambigu atau tidak jelas. Oleh karena itu, untuk penelitian ke depannya dapat melibatkan metode lain atau menggabungkan beberapa metode untuk meningkatkan akurasi dan keandalan analisis sentimen ulasan film ini.

Kata Kunci: Analisis Sentimen, IMDb, Naive Bayes, Ulasan Film

PENDAHULUAN

Di era digital saat ini, ulasan telah menjadi sumber informasi yang berharga dalam menilai produk dan layanan, termasuk dalam hal film. Ulasan film telah menjadi sumber informasi penting bagi orang-orang salah satunya yakni untuk memutuskan apakah mereka akan menonton sebuah film atau tidak.

Internet Movie Database (IMDb) adalah platform *online* yang menyajikan informasi tentang film, televisi, dan tokoh terkenal. Pengguna IMDb dapat memberi peringkat film dan merekomendasikan film menarik kepada sesama pengguna. Beberapa penelitian telah menggunakan IMDb sebagai sumber data, termasuk penelitian yang fokus pada peningkatan rekomendasi film IMDb dengan pengaturan dan filter interaktif, pengembangan sistem rekomendasi film IMDb berdasarkan konten, serta analisis sentimen masyarakat berdasarkan ulasan yang ada di platform IMDb. Selain itu, IMDb juga digunakan sebagai sumber data untuk menghasilkan *dataset* evaluasi dalam perbandingan yang dilakukan oleh penonton [1] - [4].

Analisis sentimen adalah pemrosesan bahasa alami, analisis teks, dan linguistik komputasional untuk mengidentifikasi dan mengekstrak informasi subjektif dari sumber teks. Salah satu objek yang dapat dipilih untuk melakukan analisis sentimen yaitu ulasan film. Naive Bayes adalah salah satu metode klasifikasi yang dapat digunakan secara luas dalam analisis sentimen. Metode ini berdasarkan pada Teorema Bayes dan menggunakan asumsi "naif" bahwa semua ulasan saling independen. Selain itu, algoritma Naive Bayes juga telah terbukti efektif dalam mengklasifikasikan teks dan ulasan dalam berbagai konteks. Contohnya, beberapa studi telah menggunakan metode Naive Bayes dalam menganalisis sentimen, seperti dalam analisis sentimen pada platform Twitter dan dalam mengevaluasi sentimen ulasan film yang berasal dari situs Rotten Tomatoes. Seiring dengan itu, ada juga penelitian lain yang mengadopsi pendekatan pembelajaran mesin untuk menganalisis sentimen ulasan film dan mengklasifikasikan ulasan-ulasan tersebut ke dalam berbagai kategori dengan tujuan meningkatkan akurasi analisis sentimen [5] - [8].

Penelitian ini dimaksudkan untuk mengevaluasi tingkat efektivitas metode Naive Bayes dalam mengklasifikasikan ulasan ke dalam kategori positif dan negatif. Klasifikasi tersebut berguna dalam menyusun ulasan menjadi kelompok-kelompok yang sesuai, sehingga mempermudah pemahaman.

Fery Anuar Ramadhan P, Frido Firman F, Ultach Enri

Dalam melakukan analisis sentimen terhadap ulasan film, terdapat beberapa tantangan yang perlu dihadapi, seperti penggunaan bahasa dan kalimat dalam ulasan tersebut. Oleh karena itu, diperlukan teknik yang tepat dan efektif untuk mengatasi tantangan tersebut untuk memperoleh hasil analisis sentimen yang akurat dan konsisten.

TINJAUAN PUSTAKA

Implementation of Sentiment Analysis Movie Review based on IMDB with Naive Bayes Using Information Gain on Feature Selection

Dalam penelitian tersebut, mereka menggunakan metode Naive Bayes dengan TF-IDF untuk mengambil fitur dan *Information Gain* untuk memilih fitur dalam memilih ulasan film di IMDb berdasarkan sentimen. Hasilnya menunjukkan bahwa presisi maksimal mencapai 84,50%, akurasi 88,27%, recall 88,27%, dan skor f1 sebesar 86,34%. Artinya, metodenya cukup bagus dalam mengklasifikasikan ulasan film [9].

Sentiment Analysis on Movie Review using Naive Bayes

Penelitian ini menggunakan metode Naive Bayes dengan pendekatan Bag of Words (BoW) dan pemodelan TF-IDF untuk mengelompokkan ulasan film di IMDb berdasarkan sentimen. Hasilnya menunjukkan tingkat akurasi mencapai 89% [10].

Classification of Movie Review Sentiment Analysis Using Chi-Square and Multinomial Naive Bayes with Adaptive Boosting

Penelitian selanjutnya menggunakan metode Naive Bayes multinomial dengan pemilihan fitur menggunakan chi-square dan Adaboost untuk mengelompokkan ulasan film di IMDb berdasarkan sentimen. Hasilnya menunjukkan akurasi sebesar 87,74% [11].

Sentiment Analysis Tools for Movie Review Evaluation - A Survei

Penelitian lain membandingkan performa beragam alat analisis sentimen berdasarkan leksikon, seperti Analisis Sentimen Naive Bayes Textblob, dalam mengidentifikasi sentimen dalam ulasan film. Hasilnya menunjukkan bahwa Analisis Sentimen Naive Bayes Textblob memberikan hasil terbaik dengan tingkat akurasi 73% dan skor F1 sebesar 0,78 untuk ulasan film di IMDb [12].

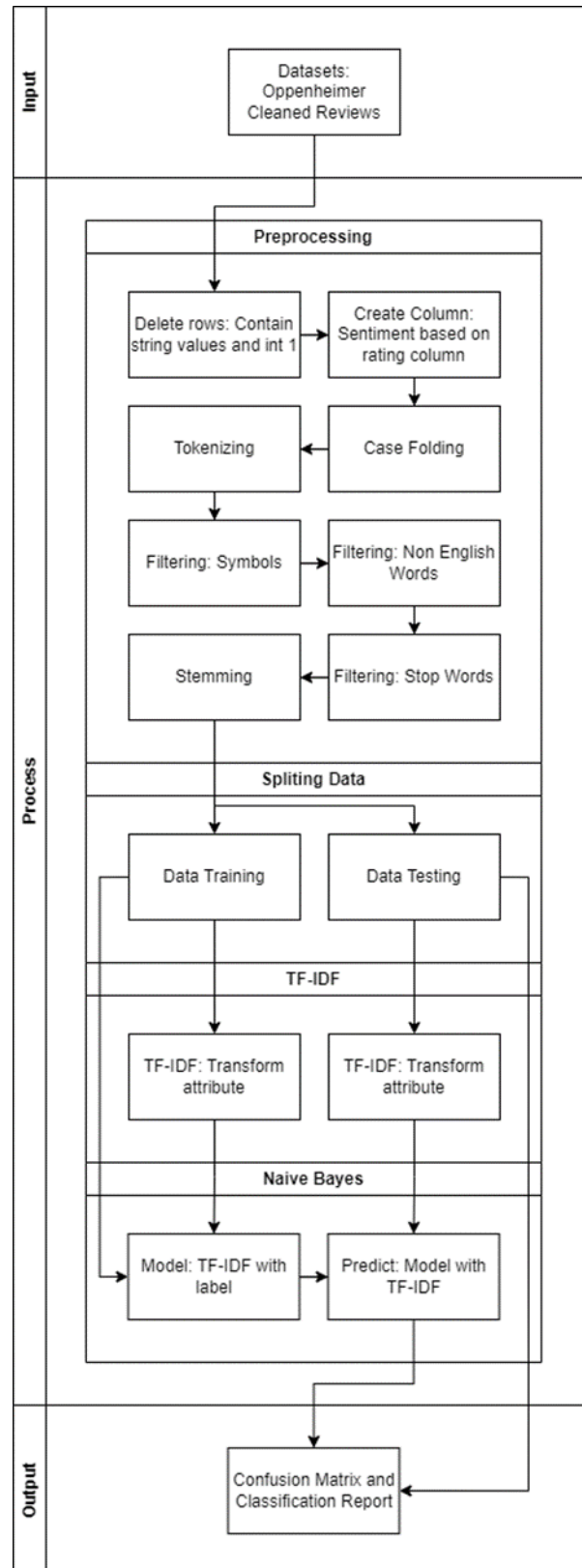
METODE

Metode Penelitian

Dalam penelitian ini, metode kuantitatif digunakan untuk mengevaluasi sentimen yang terkandung dalam ulasan film "Oppenheimer" di situs IMDb. Pendekatan utama penelitian ini adalah dengan menggunakan algoritma Naive Bayes untuk otomatis mengklasifikasikan sentimen dalam ulasan-ulasan tersebut, berfokus pada data yang dapat diukur secara numerik seperti persentase sentimen positif dan negatif.

Kerangka Penelitian

Kerangka penelitian yang dipergunakan untuk mengatasi permasalahan ini adalah metode klasifikasi dengan menggunakan algoritma Naive Bayes. Metode ini dipilih dan dimasukkan ke dalam kerangka penelitian karena kemampuannya dalam mengenali serta mengelompokkan data berdasarkan informasi yang relevan serta pola yang terdapat dalam *dataset* yang telah dikumpulkan oleh peneliti. Dengan penerapan algoritma Naive Bayes, penelitian ini dapat mengoptimalkan kinerja dan akurasi dalam proses pengklasifikasian data, yang pada gilirannya memungkinkan pencapaian tujuan penelitian dengan lebih efisien dan efektif. Berikut penjelasan mengenai kerangka penelitian pada Gambar 1.



Gambar 1. Kerangka Penelitian

1. Tahapan Pengambilan Data

Data tersebut didapatkan dari ulasan pengguna film Oppenheimer di platform IMDb. Selanjutnya, data tersebut dilakukan proses pembersihan, kemudian diubah menjadi dokumen dalam format CSV.

2. Tahapan Pre-Processing Data

Pre-processing merupakan langkah awal dalam *data mining* yang berfokus pada membersihkan, mengubah, dan menyiapkan data mentah agar siap untuk tahapan selanjutnya [13]. Langkah-langkah yang diterapkan dalam tahap ini meliputi:

- Pemadatan data dengan menghapus baris yang tidak relevan dan menambahkan kolom baru.
- Penyederhanaan huruf menjadi *lowercase* (*case folding*).
- Pembagian teks menjadi *token array* (*tokenization*).
- Penyaringan simbol yang tidak relevan.
- Penyaringan kata-kata yang tidak terdaftar dalam kamus Bahasa Inggris.
- Penggunaan daftar kata penghenti (*stop words*) sebagai penyaring.
- Mengubah kata menjadi kata dasarnya (*stemming*).

3. Tahapan Splitting Data

Splitting data dalam *data mining* mengacu pada proses membagi *dataset* menjadi dua untuk keperluan pelatihan dan pengujian. Ini adalah tahap yang sangat penting dalam *data mining* karena membantu mengevaluasi kinerja model klasifikasi Naive Bayes pada data yang belum pernah dilihat sebelumnya [14].

4. Tahapan Pembobotan TF-IDF

Pembobotan TF-IDF adalah teknik yang digunakan untuk memberi nilai kepada kata-kata yang terdapat dalam bagian sebuah atribut berdasarkan seberapa sering kata tersebut muncul dalam bagian atribut tersebut dan seberapa umum kata tersebut dalam koleksi bagian atribut secara keseluruhan. Nilai kata yang dihasilkan oleh metode ini berguna untuk mengukur tingkat signifikansi atau relevansi suatu kata dalam konteks bagian atribut tersebut [15].

5. Tahapan Pemodelan Data

Multinomial Naive Bayes (MNB) Classification merupakan pendekatan klasifikasi teks yang memanfaatkan probabilitas untuk meramalkan

kategori atau kelas suatu dokumen [16]. Model ini berasal dari data pelatihan, di mana nilai atribut yang telah diwakilkan oleh pembobotan digabungkan dengan label. Setelah mendapatkan model, langkah selanjutnya adalah melakukan prediksi dengan menggunakan data pembobotan atribut yang berasal dari data pengujian.

6. Tahapan Evaluasi Data

Tahap evaluasi model merupakan tahap terakhir dalam proses pengembangan model atau algoritma. Pada tahap ini, model yang telah diperlakukan pada data pelatihan akan diujikan pada data pengujian guna mengukur kinerjanya. Evaluasi model memanfaatkan berbagai metrik evaluasi seperti akurasi, presisi, *re-call*, F1-score, dan lainnya. Hasil evaluasi model digunakan untuk menilai kelayakan penggunaan model tersebut atau apakah perlu dilakukan peningkatan [17].

HASIL DAN PEMBAHASAN

1. Hasil Pengumpulan Data

Data ulasan film Oppenheimer diperoleh dari situs IMDb menggunakan modul Python yang disebut Scrapset. Dokumen yang sudah dibersihkan memiliki kolom peringkat yang diperbandingkan dengan skala 10. *Dataset* ini mengandung 84.048 ulasan dari pengguna, namun masih mengandung data yang ambigu yang perlu disaring. Sebagai ilustrasi tentang sejumlah besar data yang digunakan, berikut ini beberapa data yang disajikan dalam Tabel 1 dari dataset.

Tabel 1. Hasil Pengumpulan Data

Text	Rating
A challenging watch to be sure, but a worthwhi...	8
A brilliantly layered examination of a man thr...	1
This movie is the bomb!zeki-420 July 2023Maste...	8
Exceptional storytelling and Genius Cinametogr...	9
An unpopular opinion from a die-hard Nolan fan...	5
Mediocre and Meanderingjojunno26 July 2023This...	M
A cold, long history lessonapereztenessa-123 J...	5

2. Hasil Pre-Processing Data

Tahapan *Pre-Preprocessing Data* terdiri dari *labeling*, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Untuk melakukan hal ini, diperlukan beberapa *library* Python terkait seperti Pandas dan NLTK.

Langkah pertama adalah menghapus baris data yang tidak relevan dan memberikan label pada *dataset*. Sebagai hasil dari langkah-langkah tersebut, beberapa data dari *dataset* yang sudah diolah disajikan dalam bentuk Tabel 2.

Tabel 2. Hasil Pelabelan Data

Text	Rating	Sentiment
A challenging watch to be sure, but a worthwhi...	8	Positive
This movie is the bomb!zeki-420 July 2023Maste...	8	Positive
Exceptional storytelling and Genius Cinametogr...	9	Positive
An unpopular opinion from a die-hard Nolan fan...	5	Negative
A cold, long history lessonapereztenessa-123 J...	5	Negative

Tahap berikutnya adalah proses *case folding*, di mana semua atribut dikonversi menjadi huruf kecil. Sebagai hasil dari langkah-langkah tersebut, beberapa data dari *dataset* ditampilkan dalam Tabel 3.

Tabel 3. Case Folding

Text	Rating	Sentiment
a challenging watch to be sure, but a worthwhi...	8	Positive
this movie is the bomb!zeki-420 july 2023maste...	8	Positive
exceptional storytelling and genius cinametogr...	9	Positive
an unpopular opinion from a die-hard nolan fan...	5	Negative
a cold, long history lessonapereztenessa-123 j...	5	Negative

Langkah selanjutnya adalah tahap *tokenizing*, di mana atribut dipecah menjadi *array* per-kata. Sebagai hasil dari prosedur tersebut, beberapa data yang terdapat dalam *dataset* disajikan dalam Tabel 4.

Tabel 4. Tokenizing

Text	Rating	Sentiment
[a, challenging, watch, to, be, sure,, but, a, worthwhi...	8	Positive
[this, movie, is, the, bomb! zeki-420, july, 2023maste...	8	Positive
[exceptional, storytelling, and, genius, cinametogr...	9	Positive
[an, unpopular, opinion, from, a, die-hard, nolan, fan...	5	Negative
[a, cold,, long, history, lessonapereztenessa-123, j...	5	Negative

Tahap selanjutnya melibatkan penyaringan simbol, kata-kata non-Bahasa Inggris, dan kata-kata penghenti (*stop words*). Sebagai hasil dari proses tersebut, beberapa data dari *dataset* ditampilkan dalam Tabel 5.

Tabel 5. Filtering

Text	Rating	Sentiment
[watch, sure, year, many, people, included, la...	8	Positive
[movie, craftsman, probably, best, blockbuster...	8	Positive
[exceptional, storytelling, genius, came, thea...	9	Positive
[unpopular, opinion, diehard, favorite, direct...	5	Negative
[cold, long, history, marketing, campaign, two...	5	Negative

Langkah berikutnya adalah tahap *stemming*, di mana kata-kata diubah menjadi bentuk kata dasarnya. Sebagai hasil dari langkah-langkah tersebut, beberapa data dari *dataset* disajikan dalam Tabel 6.

Tabel 6. Stemming

Text	Rating	Sentiment
[watch, sure, year, mani, peopl, includ, larg...	8	Positive
[movi, craftsman, probabl, best, blockbust, di...	8	Positive
[except, storytel, genui, came, theater, watch...	9	Positive
[unpopular, opinion, diehard, favorit, directo...	5	Negative
[cold, long, histori, market, campaign, two, f...	5	Negative

3. Hasil Splitting Data

Pembagian data adalah tindakan yang dilakukan untuk memisahkan total data keseluruhan, yang berjumlah 41.078 data, menjadi dua kelompok. Kelompok pertama adalah data pelatihan, yang terdiri dari 80% dari total data atau setara dengan 32.682 data. Sementara itu, kelompok kedua adalah data pengujian, yang mencakup 20% dari total data atau sebanyak 8.216 data.

4. Hasil Pembobotan TF-IDF

Pembobotan TF-IDF adalah tahap penting dalam mengevaluasi signifikansi kata-kata dalam konteks tertentu, seperti dalam dokumen. Proses ini melibatkan tiga langkah utama yaitu menghitung frekuensi kata dalam dokumen tertentu, mengukur seberapa umum kata tersebut dalam koleksi dokumen keseluruhan, dan mengalikan hasilnya untuk menghasilkan nilai TF-IDF. Ilustrasi tahapan ini divisualisasikan dalam contoh pada Tabel 7.

Tabel 7. Perhitungan TF-IDF

Kata	TF			IDF	TF * IDF		
	T1	T2	T3		T1	T2	T3
Sure	1	1	0	0.1761	0.1761	0.1761	0
Best	1	0	1	0.1761	0.1761	0	0.1761
Diehard	0	0	1	0.4771	0	0	0.4771

5. Hasil Pemodelan Data

Pada tahap pemodelan data, salah satu pustaka yang digunakan adalah Multinomial Naive Bayes dari Sklearn. Untuk melatih model, perlu menggunakan data TF-IDF dari data pelatihan serta labelnya. Setelah model berhasil dibuat, langkah selanjutnya adalah melakukan prediksi dengan menggunakan data TF-IDF dari data pengujian.

6. Hasil Evaluasi Data

Setelah berhasil melakukan prediksi, langkah selanjutnya adalah melakukan evaluasi menggunakan *Confusion Matrix* dan *Classification Report* dengan membandingkan hasil prediksi dengan label data pelatihan. Hasil perhitungan ini terdokumentasi dalam Gambar 2.

Confusion Matrix:

```
[[1160 291]
 [ 8 6757]]
```

Classification Report:

	precision	recall	f1-score	support
negative	0.99	0.80	0.89	1451
positive	0.96	1.00	0.98	6765
accuracy			0.96	8216
macro avg	0.98	0.90	0.93	8216
weighted avg	0.96	0.96	0.96	8216

Gambar 2. Hasil Evaluasi Data

Menurut Gambar 2, hasil pengujian menunjukkan tingkat akurasi sebesar 96%, nilai presisi rata-rata mencapai 98%, *re-call* rata-rata sebesar 90%, dan *f1-score* rata-rata sebesar 93%. Terdapat kesalahan dalam mengklasifikasikan sentimen, yang disebabkan oleh beberapa faktor seperti pengguna yang memberikan *rating* di bawah 6, meskipun isi ulasannya mungkin mengandung sentimen positif. Selain itu, penggunaan bahasa Inggris yang kurang tepat juga berkontribusi pada ketidakakuratan klasifikasi.

KESIMPULAN

Penelitian ini berhasil memisahkan antara sentimen positif dan negatif dengan tingkat akurasi yang tinggi menggunakan data ulasan dan peringkat yang diberikan oleh pengguna untuk film Oppenheimer di platform IMDb. Penelitian ini menggunakan total 41.078 data yang secara acak dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian. Hasil penelitian mencatat tingkat akurasi sebesar 96%, dengan nilai presisi rata-rata mencapai 98%, nilai *re-call* rata-rata sebesar 90%, dan nilai rata-rata *f1-score* sebesar 93%. Terdapat kesalahan dalam mengklasifikasikan sentimen, yang dapat disebabkan oleh berbagai faktor seperti pengguna yang memberikan peringkat di bawah 6, meskipun ulasan mereka mungkin mengandung sentimen positif. Selain itu, ketidakakuratan klasifikasi juga bisa dipengaruhi oleh penggunaan bahasa Inggris yang kurang tepat.

Temuan dari penelitian ini memberikan wawasan berharga bagi industri film dalam memahami respons pengguna terhadap film Oppenheimer. Informasi sentimen yang diperoleh dapat mendukung pengambilan keputusan yang lebih baik bagi pembuat kebijakan industri, produser, dan penonton terkait pengembangan dan pemasaran film. Untuk penelitian mendatang dapat

menjelajahi penggunaan metode lain atau menggabungkan beberapa metode guna meningkatkan akurasi dan keandalan analisis sentimen yang dilakukan.

DAFTAR PUSTAKA

- [1] S. Dooms, T. De Pessemier, dan L. Martens, "Improving IMDb Movie Recommendations with Interactive Settings and Filters," 2014. [Daring]. Tersedia pada: <http://www.imdb.com>
- [2] C. Wei, X. Chen, Z. Tang, dan W. Cheng, "Fully content-based IMDb movie recommendation engine with Pearson similarity," dalam *International Conference on Green Communication, Network, and Internet of Things (GCNIoT 2021)*, J. Mou dan S. Chen, Ed., SPIE, Des 2021, hlm. 15. doi: 10.1117/12.2624930.
- [3] S. Aji, I. Maryani, dan E. Muningsih, "Analisis Sentiment Masyarakat Menggunakan Penggabungan Algoritma Naive Bayes Dan Particle Swarm Optimization," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 7, no. 2, Des 2022, doi: 10.31294/ijcit.v7i2.14086.
- [4] N. Pavlichenko dan D. Ustalov, "IMDB-WIKI-SbS: An Evaluation Dataset for Crowdsourced Pairwise Comparisons," Okt 2021.
- [5] A. K. K. P, L. Celestine S, dan V. V Kumar, "Naive Bayes Algorithm for Sentiment Analysis on Twitter," dalam *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, IEEE, Jul 2021, hlm. 1-4. doi: 10.1109/ICSCAN53069.2021.9526473.
- [6] S. Rizal, Adiwijaya, dan M. D. Purbolaksono, "Sentiment Analysis on Movie Review from Rotten Tomatoes Using Word2Vec and Naive Bayes," dalam *2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT)*, IEEE, Nov 2022, hlm. 180-185. doi: 10.1109/ICoSEIT55604.2022.10030009.
- [7] Md. T. Zumma, K. B. M. Tahmiduzzaman, R. Khan, N. S. Roni, A. A. M. Rahat-Bin-Rafique, dan A. H. Akash, "Sentimental Analysis of Movie Review using Machine Learning Approach," dalam *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*, IEEE, Des 2022, hlm. 1-5. doi: 10.1109/CCET56606.2022.10080860.
- [8] D. Vishnu, "Multi Class Data Classification to Improve Accuracy in Sentiment Analysis using Machine Learning," *Int J Res Appl Sci Eng Technol*, vol. 9, no. VI, hlm. 1457-1461, Jun 2021, doi: 10.22214/ijraset.2021.35291.
- [9] M. Mahyarani, A. Adiwijaya, S. Al Faraby, dan M. Dwifabri, "Implementation of Sentiment Analysis Movie Review based on IMDB with Naive Bayes Using Information Gain on Feature Selection," dalam *2021 3rd International Conference on Electronics Representation and Algorithm (ICERA)*, IEEE, Jul 2021, hlm. 99-103. doi: 10.1109/ICERA53111.2021.9538763.
- [10] N. L. Adam, N. H. Rosli, dan S. C. Soh, "Sentiment Analysis on Movie Review using Naïve Bayes," dalam *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, IEEE, Sep 2021, hlm. 1-6. doi: 10.1109/AiDAS53897.2021.9574419.
- [11] M. B. Hamzah, "Classification of Movie Review Sentiment Analysis Using Chi-Square and Multinomial Naïve Bayes with Adaptive Boosting," *Journal of Advances in Information Systems and Technology*, vol. 3, no. 1, hlm. 67-74, Apr 2021, doi: 10.15294/jaist.v3i1.49098.
- [12] M. Tetteh dan M. Thushara, "Sentiment Analysis Tools for Movie Review Evaluation - A Survey," dalam *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, Mei 2023, hlm. 816-823. doi: 10.1109/ICICCS56967.2023.10142834.
- [13] H. Henderi dan R. L. Wanda, "PREPROCESSING DATA UNTUK SISTEM PERAMALAN TINGKAT KEDISIPLINAN MAHASISWA," *ICIT Journal*, vol. 3, no. 2, hlm. 296-308, Agu 2017, doi: 10.33050/icit.v3i2.70.
- [14] B. L. Aurelian, "An information entropy based splitting criterion better for the Data Mining Decision Tree algorithms," dalam *2018 22nd International Conference on System Theory, Control and Computing (ICSTCC)*, IEEE, Okt 2018, hlm. 535-540. doi: 10.1109/ICSTCC.2018.8540743.

- [15] L. Mayasari dan D. Indarti, "KLASIFIKASI TOPIK TWEET MENGENAI COVID MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES DENGAN PEMBOBOTAN TF-IDF," *Jurnal Ilmiah Informatika Komputer*, vol. 27, no. 1, hlm. 43-53, Apr 2022, doi: 10.35760/ik.2022.v27i1.6184.
- [16] G. Subroto, N. Sulistiyowati, dan A. A. Ridha, "Klasifikasi Jenis Kekerasan Pada Perempuan Dan Anak Dengan Algoritma Multinomial Naïve Bayes," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 5, no. 1, hlm. 104-113, Jun 2022, doi: 10.31539/intecom.s.v5i1.3598.
- [17] M. Z. Fuadi, Y. Gloriani, dan D. E. Mascita, "MODEL SINEKTIK DENGAN MEDIA AUDIOVISUAL DALAM PEMBELAJARAN TEKS CERITA INSPIRATIF," *JURNAL TUTURAN*, vol. 11, no. 2, hlm. 89, Des 2022, doi: 10.33603/jt.v11i2.7693.