



# Implementation of Web-Based Sentiment Analysis Application on Movie Reviews Using the Naive Bayes Algorithm

Ismaturrahmi<sup>1</sup>, M. Khadafi<sup>2\*</sup>, Amirullah<sup>3</sup>

<sup>1,2,3</sup> Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe Jln. B.Aceh Medan Km.280 Buketrata 24301  
INDONESIA

\*Penulis Korespondensi : mkhadafi@pnl.ac.id

## INFORMASI ARTIKEL

*Riwayat artikel:*  
Diajukan pada 18 Mei 25  
Direvisi pada 03 Juni 25  
Publikasi pada 20 Juni 25

*Kata kunci:*  
Ulasan  
Naive Bayes  
Analisis Sentimen  
Klasifikasi

*Keywords:*  
Review  
Naive Bayes  
Sentiment Analysis  
Classification

## ABSTRAK

Kemajuan teknologi telah membawa perubahan signifikan dalam industri film, termasuk kemudahan akses ke berbagai platform ulasan dan penilaian konten seperti IMDb. Peningkatan jumlah penonton film telah menyebabkan lonjakan jumlah ulasan film, sehingga analisis sentimen menjadi penting untuk memahami reaksi penonton terhadap suatu film. Analisis sentimen menggunakan teknologi untuk mengevaluasi dan memahami ekspresi emosional dalam ulasan film, membantu mengidentifikasi sentimen positif, netral, atau negatif. Penelitian ini bertujuan agar pengguna tidak perlu lagi membaca ulasan panjang yang membutuhkan waktu lama sehingga dapat membantu dalam membuat keputusan yang lebih baik terkait film yang ingin ditonton. Dengan adanya sistem ini, pengguna bisa langsung mengetahui sentimen ulasan tanpa harus membaca secara menyeluruh menggunakan algoritma Naive Bayes. Data ulasan film yang digunakan dalam penelitian ini diambil dari platform Kaggle, terdiri dari 1000 data, di mana 800 data digunakan untuk pelatihan dan 200 data untuk pengujian. Hasil pengujian menunjukkan bahwa sistem berhasil diimplementasikan dengan baik. Model Naive Bayes yang digunakan menunjukkan kinerja yang memadai dengan akurasi 81%. Angka ini diperoleh melalui evaluasi menggunakan *confusion matrix*, yang mengukur kinerja model dalam mengklasifikasikan ulasan ke dalam kategori sentimen positif, netral, dan negatif. *Confusion matrix* menunjukkan bahwa model ini cukup efektif dalam mengenali sentimen positif, namun kurang akurat dalam mengidentifikasi sentimen negatif dan netral. Hal ini menandakan model Naive Bayes mampu memberikan gambaran umum yang baik tentang sentimen positif dalam ulasan. Penelitian ini diharapkan dapat memberikan wawasan yang lebih baik kepada penonton mengenai sentimen ulasan film.

## ABSTRACT

*Technological advances have brought significant changes to the film industry, including easy access to various content review and rating platforms such as IMDb. The increasing number of moviegoers has led to a surge in the number of movie reviews, making sentiment analysis important to understand audience reactions to a movie. Sentiment analysis uses technology to evaluate and understand emotional expressions in movie reviews, helping to identify positive, neutral, or negative sentiments. This study aims to eliminate the need for users to read long reviews that take a long time. With this system, users can immediately find out the sentiment of the review without having to read it all using the Naive Bayes algorithm. The movie review data used in this study was taken from the Kaggle platform, consisting of*

---

*1000 data, where 800 data were used for training and 200 data for testing. The test results showed that the system was successfully implemented. The Naïve Bayes model used showed adequate performance with an accuracy of 81%. This figure was obtained through an evaluation using a confusion matrix, which measures the model's performance in classifying reviews into positive, neutral, and negative sentiment categories. The confusion matrix shows that this model is quite effective in recognizing positive sentiment, but less accurate in identifying negative and neutral sentiments. This indicates that although the Naïve Bayes model is able to provide a good overview of positive sentiment in reviews, there is room for improvement especially in the classification of negative and neutral sentiments. This study is expected to provide better insight to the audience regarding the sentiment of movie reviews, thus helping in making better decisions regarding the movies they want to watch.*

---

## 1. Pendahuluan

Kemajuan teknologi saat ini telah membawa perubahan besar dalam industri film, dengan banyak platform penjualan dan distribusi konten yang menyediakan akses mudah ke berbagai film dan serial TV. Dalam beberapa tahun terakhir, jumlah penonton film meningkat secara drastis yang juga menyebabkan peningkatan jumlah ulasan film. Film merupakan seni yang sangat populer saat ini dengan beragam jenis dan penggemarnya masing-masing. Setiap penggemar memberikan penilaian melalui *rating* dan ulasan yang mencakup berbagai aspek seperti cerita, akting, dan efek visual. Ulasan ini mengandung sentimen, baik positif maupun negatif. Analisis sentimen dapat membantu dalam mengelompokkan ulasan film. Salah satu metode yang bisa digunakan untuk analisis sentimen adalah klasifikasi Naïve Bayes [1].

Klasifikasi adalah algoritma yang memanfaatkan data dengan target (kelas/label) yang berbentuk nilai kategorial atau nominal. Sebagai contoh, klasifikasi dapat digunakan untuk menganalisis apakah sebuah ulasan film bersifat positif atau negatif [2]. Ulasan film penting bagi semua orang untuk memperoleh informasi mengenai sebuah film, sehingga dapat membantu mereka memahami konten film yang akan ditonton [3]. Dengan meningkatnya jumlah ulasan film, penggunaan analisis sentimen menjadi penting dalam memahami reaksi penonton terhadap suatu film. Analisis sentimen melibatkan penggunaan teknologi untuk mengevaluasi dan memahami ekspresi emosional dalam ulasan-ulasan tersebut. Analisis sentimen adalah teknik ekstraksi informasi dari teks dengan tujuan memahami sentimen, apakah itu positif, netral, atau negatif. Pengguna internet memanfaatkan metode ini di platform media sosial untuk menyampaikan penilaian mereka. Analisis sentimen dapat digunakan untuk mengekstrak data opini, secara otomatis memahami, dan memproses data teks untuk mengidentifikasi sentimen yang terkandung dalam suatu opini [4].

Teks ulasan adalah tulisan yang dihasilkan melalui kegiatan mengupas, menafsirkan, atau mengapresiasi suatu karya tertentu, seperti penayangan film [5]. Salah satu platform yang dapat digunakan untuk melihat dan memberikan komentar tentang film adalah IMDb. IMDb merupakan situs web yang bermanfaat untuk mengetahui berbagai detail tentang sebuah film, seperti aktor/aktris yang berperan, sinopsis film, *rating*, dan ulasan film [6]. Ulasan adalah bentuk penilaian seseorang terhadap produk atau layanan. Ulasan dapat menjadi sumber informasi penting bagi pelanggan lain yang ingin mengetahui kualitas produk atau layanan tersebut [7].

Penggunaan analisis sentimen pada ulasan pengguna bertujuan untuk menyelidiki ulasan produk dalam menentukan opini atau perasaan terhadap suatu film secara keseluruhan [8]. Algoritma Naive Bayes

merupakan salah satu algoritma *supervised learning* yang dapat digunakan untuk klasifikasi. Metode ini bekerja dengan cepat dan efisien dalam mengklasifikasikan teks [9].

Naive Bayes adalah algoritma *data mining* yang mudah digunakan dengan waktu pemrosesan cepat. Algoritma ini efektif karena strukturnya yang sederhana dan cocok untuk memprediksi probabilitas kelas. Berdasarkan teorema Bayes, metode ini memiliki kinerja serupa dengan *decision tree* dan *neural network*, serta terbukti akurat dan cepat, terutama pada besar [10].

Penelitian terdahulu Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi mendapatkan akurasi 94% tanpa penambahan fitur [4]. Selanjutnya penelitian Analisis Sentimen Review Film Menggunakan Naive Bayes Classifier Dengan Fitur TF-IDF mendapatkan akurasi 86.48% [9].

Oleh karena itu, penggunaan algoritma Naive Bayes pada penelitian ini mengenai analisis sentimen pada ulasan film, diharapkan dapat memberikan wawasan yang lebih baik kepada penonton tentang suatu film dan pemahaman mendalam terhadap sentimen ulasan film diharapkan dapat menjadi pedoman bagi penonton dalam membuat keputusan tentang film yang ingin mereka tonton.

## 2. Metode

### 2.1 Data dan Pengumpulan Data

Pengumpulan data untuk penelitian ini dilakukan melalui website Kaggle dengan melakukan pencarian menggunakan kata kunci "IMDb Review". Setelah melakukan pencarian, *dataset* yang relevan dengan ulasan pengguna di platform IMDb telah diidentifikasi. *Dataset* tersebut berisi informasi terkait ulasan yang diberikan oleh pengguna untuk berbagai film yang ada di IMDb. Data film yang akan diambil berupa The Shawshank Redemption, The Godfather, Spider-Man: Across the Spider-Verse, The Dark Knight, Schindler's List.

### 2.2 Analisis Kebutuhan Fungsional

Ketentuan fungsional meliputi setiap langkah yang dijalankan oleh entitas terkait, yang dapat mendukung kelancaran proses dalam sistem.

#### 1. Kebutuhan Fungsional Admin

Kebutuhan fungsional admin mencakup *login* dan *logout* untuk keamanan akses sistem. Admin dapat mengelola data film, termasuk penambahan, pengeditan, dan penghapusan informasi film. Selain itu, admin bertanggung jawab untuk mengelola ulasan pengguna, serta mengatur data latih dan uji yang digunakan dalam pelatihan model. Admin juga dapat melatih dan menguji model analisis sentimen, serta melihat hasil analisis,

#### 2. Kebutuhan Fungsional Pengguna

Melihat film dan melihat hasil analisis sentimen. Kebutuhan fungsional pengguna mencakup kemampuan untuk melihat daftar film yang tersedia. Pengguna juga dapat mengakses hasil analisis sentimen dari ulasan film, yang menunjukkan kategori sentimen positif, negatif, atau netral.

### 2.3 Analisis Kebutuhan Non Fungsional

Kebutuhan non-fungsional mencakup persyaratan perangkat lunak dan perangkat keras yang diperlukan untuk menjalankan sistem. Kebutuhan non-fungsional meliputi:

1. Perangkat Lunak (*software*):

Windows 11, Google Chrome, Visual Studio Code, Laragon (MySQL, PHP), dan Python.

2. Perangkat Keras (*hardware*)

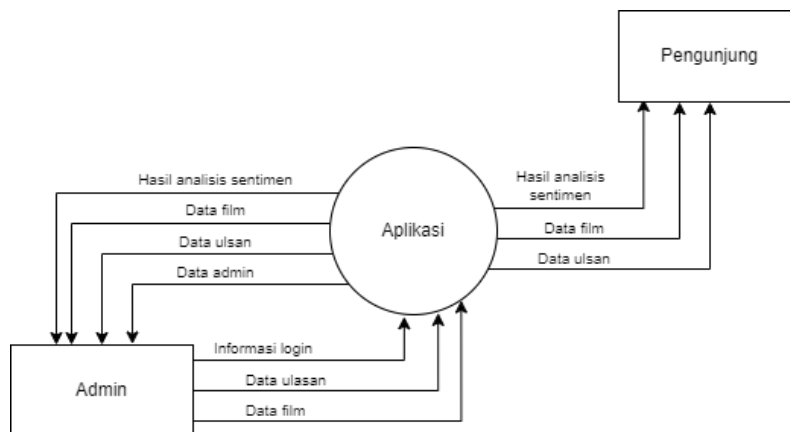
Processor AMD Ryzen 3 atau Intel i3 gen 6, RAM 4GB, dan Storage 100GB

### 2.4 Rancangan Sistem (*software/hardware*)

*Data Flow Diagram* (DFD) digunakan dalam proses perancangan untuk membangun sistem ini. Perancangan sistem ini melibatkan penggunaan *context diagram*, alur perancangan model, dan alur *labelling data*.

#### 2.4.1 Context Diagram

Pada sistem ini, terdapat dua jenis user yang berbeda, yaitu admin dan pengunjung. Setiap user memiliki tingkat akses yang berbeda, termasuk hak untuk membaca, membuat, memperbarui, dan menghapus. Adapun *context diagram* dapat dilihat pada Gambar 1.

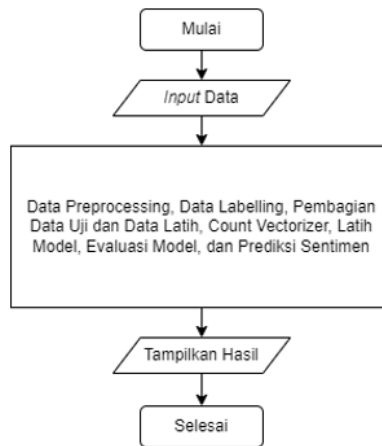


**Gambar 1.** Context Diagram

Gambar 1 menunjukkan *context diagram* sistem aplikasi yang terdapat dua entitas utama dalam diagram ini pengunjung dan admin. Aplikasi berperan sebagai pusat pengolahan data, menerima *input* dari admin dan memberikan *output* kepada pengunjung dan admin. Admin memiliki peran penting dalam sistem ini, dapat memasukkan informasi *login*, data ulasan, dan data film. Sebagai balasan, Aplikasi menyediakan admin hasil analisis sentimen, data film, data ulasan, dan data admin. Sedangkan pengunjung, hanya menerima *output* dari aplikasi berupa hasil analisis sentimen, data film, dan data ulasan. Tidak ada aliran data dari pengunjung ke aplikasi, yang mengindikasikan bahwa pengunjung dalam sistem ini terbatas pada membaca informasi yang disediakan.

#### 2.4.2 Alur Perancangan Model

Alur perancangan model Naive Bayes menjelaskan langkah-langkah utama, mulai dari *input* data hingga hasil. Setiap tahap penting untuk memastikan model memberikan prediksi akurat dan andal. Gambar 2 menunjukkan hubungan antar komponen dalam proses perancangan, memudahkan pemahaman keseluruhan.



**Gambar 2.** Alur Perancangan Model

Gambar 2 menunjukkan alur pengembangan sistem Naive Bayes yang terdiri dari beberapa langkah, yaitu memulai alur, *input data*, *preprocessing* teks, pelabelan sentimen, pembagian data menjadi set pelatihan dan pengujian, mengubah teks menjadi vektor, melatih model, evaluasi model, prediksi, dan menampilkan hasil. Alur ini berakhir dengan penyajian hasil prediksi dalam format yang mudah dipahami. Berikut penjelasan setiap langkah dalam alur pada Gambar 2.

1. *Input Data*

Pada tahap ini, data ulasan film dimasukkan ke dalam model untuk diproses lebih lanjut. Data yang dimasukkan mencakup variabel seperti nama film dan ulasan.

2. *Data Preprocessing*

Sebelum data ulasan diproses oleh model, dilakukan tahap *preprocessing* untuk membersihkan teks agar lebih mudah dianalisis. Proses ini mencakup penghapusan karakter khusus seperti tanda baca, angka, atau simbol yang tidak relevan dengan analisis sentimen. Teks juga diubah menjadi huruf kecil agar seragam, sehingga tidak ada perbedaan dalam pengolahan kata-kata dengan kapitalisasi berbeda. *Stopwords*, yaitu kata-kata umum yang tidak memberikan makna signifikan dihapus untuk meningkatkan fokus pada kata-kata penting. Selain itu, dilakukan *stemming* untuk mengubah kata-kata menjadi bentuk dasar.

3. *Data Labelling*

Pada tahap ini, setiap ulasan akan melalui proses pelabelan sentimen. Tahap ini menentukan sentimen yang terkandung di dalamnya, apakah ulasan tersebut positif, negatif, atau netral.

4. Pembagian Data

Setelah proses pelabelan, data ulasan dibagi menjadi dua set: set pelatihan 80% dan set pengujian 20%. 80% dari data digunakan sebagai set pelatihan, yang berfungsi untuk melatih model Naive Bayes agar mampu mengenali pola sentimen dalam ulasan. Sisanya, 20%, digunakan sebagai set pengujian untuk mengevaluasi kinerja model setelah dilatih.

5. *Count Vectorizer*

Setelah proses labelling, tahap selanjutnya adalah menggunakan *Count Vectorizer* untuk mengubah teks ulasan menjadi vektor angka. Metode ini menghitung frekuensi kemunculan setiap kata dalam ulasan dan merepresentasikannya dalam bentuk vektor. Proses ini penting karena model *machine learning*, seperti Naive

Bayes, memerlukan *input* dalam format numerik untuk melakukan analisis. Hasil dari proses ini akan digunakan dalam tahap pelatihan model untuk meningkatkan akurasi prediksi sentimen.

#### 6. Latih Model

Pada tahap ini, algoritma Naive Bayes dilatih menggunakan data latih yang telah disiapkan sebelumnya. Proses pelatihan ini melibatkan perhitungan probabilitas setiap kata yang terdapat dalam teks ulasan untuk masing-masing kelas sentimen, yaitu positif, negatif, dan netral. Hasil dari proses pelatihan ini adalah model yang siap digunakan untuk memprediksi sentimen pada ulasan baru yang belum pernah dianalisis sebelumnya, dengan harapan dapat memberikan akurasi yang tinggi dalam klasifikasi sentimen.

#### 7. Evaluasi Model

Setelah model dilatih, langkah selanjutnya adalah mengevaluasi akurasinya menggunakan data uji yang telah dipisahkan sebelumnya. Proses evaluasi ini penting untuk menilai seberapa baik model dapat memprediksi sentimen pada ulasan yang belum pernah dilihatnya. Metrik yang digunakan dalam evaluasi meliputi akurasi, presisi, dan *recall*. Akurasi mengukur seberapa banyak prediksi yang benar dibandingkan dengan total prediksi, sementara presisi menilai proporsi prediksi positif yang benar di antara semua prediksi positif. *Recall*, di sisi lain, mengukur kemampuan model dalam menemukan semua contoh positif yang ada dalam data uji.

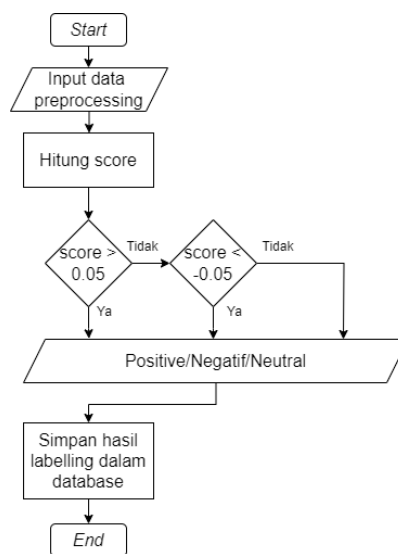
#### 8. Prediksi

Pada tahap ini, model yang telah dilatih dan dievaluasi diaplikasikan untuk memprediksi sentimen pada teks baru yang masuk. Dengan menggunakan model yang telah dilatih, kita dapat mengklasifikasikan teks baru ke dalam kategori sentimen positif, negatif, atau netral berdasarkan probabilitas yang dihitung.

#### 9. Tampilkan Hasil

Setelah proses prediksi selesai, langkah selanjutnya adalah menampilkan hasil prediksi dalam format yang mudah dipahami. Hasil ini disajikan dalam bentuk grafik yang menunjukkan ulasan beserta label sentimen yang diprediksi.

#### 2.4.3 Alur *Labelling Data*



**Gambar 3.** Alur *Labelling Data*

Gambar 3 menunjukkan alur *labelling* data. Alur ini terdiri dari beberapa langkah, mulai dari *start* hingga *end*. Dimulai dengan langkah *start*, proses ini diawali dengan data yang telah diproses sebelumnya dan dimasukkan ke dalam sistem. Kemudian, dilakukan perhitungan skor sentimen dari teks, yang menentukan apakah teks memiliki sentimen positif, negatif, atau netral. Jika skor lebih besar atau sama dengan 0.05, teks diberi label positif. Jika skor kurang dari atau sama dengan -0.05, teks diberi label negatif. Jika skor berada di antara -0.05 dan 0.05, teks diberi label netral. Hasil dari proses *labelling* ini kemudian disimpan untuk analisis lebih lanjut, dan proses diakhiri dengan langkah *end*.

### 3. Hasil Dan Pembahasan

#### 3.1 Implementasi Proses

Implementasi proses menjelaskan langkah-langkah dan prosedur yang dilakukan berdasarkan rancangan model.

##### 3.1.1 *Input Data*

Bagian ini menerima data ulasan yang mencakup variabel seperti nama film dan ulasan. Variabel-variabel ini akan diproses lebih lanjut untuk menghasilkan *output* berupa klasifikasi sentimen menggunakan metode Naive Bayes, yang membantu dalam menentukan sentimen dari setiap ulasan.

| No | Nama Film                | Ulasan   |
|----|--------------------------|--|
| 1  | The Shawshank Redemption | Three words: "Cool Hand Luke." Same film, done better, done earlier. For that matter, is this film any better than other Steven King "novelettes" such as "Stand By Me"? All in all, it probably ranks a 6 or a 7, but since people on this site have lost their minds as regards this film, I give it a 1 in one man's attempt at sanity. |
| 2  | The Shawshank Redemption | Another one of those overlong morally right-on movies that never rises above average - saved only by the acting. Its bleeding-heart politics 'dignity of individual maintained in adverse conditions' is played-out well enough but actually lacks any real  |

**Gambar 4.** *Input Data*

##### 3.1.2 *Input Data*

Pada bagian ini, data yang sudah diunggah akan dilakukan *preprocessing*. Hasil ditunjukkan pada gambar 5. Gambar 5 menunjukkan hasil *preprocessing*. Data tersebut dilakukan pembersihan seperti menghapus emoji, setelah itu dilakukan *tokenizing* dan *casefolding*, kemudian dilakukan *stopword removal* dan terakhir dilakukan *stemming*.

| Cleaned Text   | Tokens Casefolding   | Stopword Text  | Stemming Text  |
|--|--|--|--|
| three word cool hand<br>luke film done better<br>done earlier matter film<br>better steven king<br>novelette stand<br>probably rank since<br>people site lost mind<br>regard film give one<br>man attempt sanity | three word cool hand<br>luke film done better<br>done earlier matter film<br>better steven king<br>novelette stand<br>probably rank since<br>people site lost mind<br>regard film give one<br>man attempt sanity | three word cool hand<br>luke film done better<br>done earlier matter film<br>better steven king<br>novelette stand<br>probably rank since<br>people site lost mind<br>regard film give one<br>man attempt sanity | three word cool hand<br>luke film done better<br>done earlier matter<br>film better steven king<br>novelett stand<br>probabl rank sinc<br>peopl site lost mind<br>regard film give one<br>man attempt saniti |
| another one overlong<br>morally right movie<br>never rise average saved<br>acting bleeding heart<br>politics dignity<br>individual maintained<br>adverse condition<br>played well enough                         | another one overlong<br>morally right movie<br>never rise average saved<br>acting bleeding heart<br>politics dignity<br>individual maintained<br>adverse condition<br>played well enough                         | another one overlong<br>morally right movie<br>never rise average saved<br>acting bleeding heart<br>politics dignity<br>individual maintained<br>adverse condition<br>played well enough                         | anoth one overlong<br>moral right movi<br>never rise averag save<br>act bleed heart polit<br>digniti individu<br>maintain advers<br>condit play well<br>enough actual lack                                   |

**Gambar 5.** Hasil *Preprocessing*

### 3.1.3 Hasil *Labelling*

Pada bagian ini, data yang sudah dilakukan *preprocessing* akan dilakukan dilakukan *labelling* antara positif, negatif, atau netral. Gambar 6 menunjukkan hasil *labelling* salah satu ulasan. Hasil *labelling* menunjukkan positif. Data tersebut dilakukan *labelling* pada setiap ulasan, hasil dari *labelling* tersebut berupa positif, negatif, atau netral.

**Score      Sentiment**

---

0.8442      positive

**Gambar 6.** Hasil *Labelling*

### 3.1.4 Hasil Implementasi Metode Naive Bayes

Pada bagian ini, dilakukan analisis sentimen menggunakan Naive Bayes. Gambar 7 menunjukkan hasil implementasi metode Naive Bayes. Hasil implementasi ini menunjukkan salah satu hasil sentimen ulasan. *Predicted sentiment* merujuk pada hasil dari proses analisis sentimen yang dilakukan oleh model atau algoritma. Disini pertama dilakukan pembagian data latih 70 % dan data uji 30%. Kemudian dilakukan proses mengubah teks menjadi vektor angka berdasarkan frekuensi kata. Setelah itu dilakukan proses latih model dan prediksi menggunakan metode Naive Bayes.

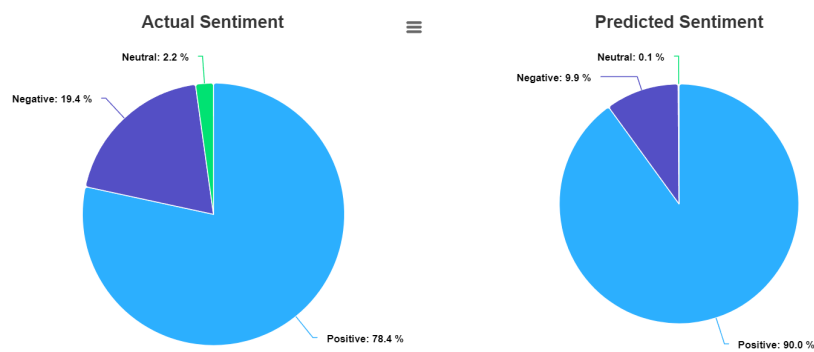
**Predicted  
Sentiment**

---

positive

**Gambar 7.** Hasil Implementasi Metode Naive Bayes

Gambar 8 menunjukkan hasil *actual* dan *prediction*. Terdapat 2 buah diagram lingkaran, yaitu *Actual Sentiment* dan *Predicted Sentiment*. Pada *Actual Sentiment*, jumlah data positif memiliki persentase sebanyak 78.4%, jumlah data negatif memiliki persentase sebanyak 19.4%, dan jumlah data netral memiliki persentase sebanyak 2.2%. Sedangkan untuk *Predicted Sentiment*, data positif memiliki persentase sebanyak 90.0%, data negatif memiliki persentase 9.9%, dan data netral memiliki persentase sebanyak 0.1%.



**Gambar 8.** Actual dan Prediction

Hasil klasifikasi menunjukkan adanya perbedaan antara sentimen aktual dan prediksi. Model Naive Bayes cenderung memprediksi sentimen positif lebih tinggi dibandingkan data aktual, dengan selisih 11.6%. Sebaliknya, terjadi *underprediction* pada sentimen negatif dan netral. Sentimen negatif diprediksi 9.5% lebih rendah dari aktual, sementara sentimen netral mengalami penurunan sebesar 2.1% dalam prediksi.

### 3.1.5 Hasil Pengujian *Confusion Matrix*

Pada bagian ini, dilakukan pengujian pada hasil analisis sentimen menggunakan metode Naive Bayes. Hasil ditunjukkan sebagai berikut.

```

Accuracy Score = 0.8080808080808081

Confusion Matrix =
[[ 5  0 28]
 [ 0  0  9]
 [ 1  0 155]]

Classification Report =

```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.83      | 0.15   | 0.26     | 33      |
| neutral      | 0.00      | 0.00   | 0.00     | 9       |
| positive     | 0.81      | 0.99   | 0.89     | 156     |
| accuracy     |           |        | 0.81     | 198     |
| macro avg    | 0.55      | 0.38   | 0.38     | 198     |
| weighted avg | 0.77      | 0.81   | 0.74     | 198     |

**Gambar 9.** Hasil Pengujian *Confusion Matrix*

Gambar 9 menunjukkan hasil pengujian *confusion matrix* dan klasifikasi pada analisis sentimen. Model ini mencapai akurasi keseluruhan 0.81 atau 81%. Kinerja model sangat baik pada sentimen positif, dengan *recall* tinggi sebesar 0.99 dan *precision* 0.81. Namun, model mengalami kesulitan dalam mendeteksi sentimen negatif dengan *recall* hanya 0.15 dan *precision* 0.83. Model gagal total dalam mengklasifikasikan sentimen netral, dengan semua metrik bernilai 0.00. Ketidakseimbangan kelas dalam *dataset* terlihat jelas, dengan mayoritas sampel positif (156), sedikit negatif (33), dan sangat sedikit netral (9). Ini mungkin berkontribusi pada kinerja yang tidak merata antar kelas. F1-score untuk kelas positif cukup tinggi (0.89), sedangkan untuk kelas negatif rendah (0.26), dan kelas netral 0.00. Nilai *macro average* dan *weighted average* juga mencerminkan ketidakseimbangan ini, dengan *macro average* yang lebih rendah (*precision* 0.55, *recall* 0.38, *f1-score* 0.38) dibandingkan *weighted average* (*precision* 0.77, *recall* 0.81, *f1-score* 0.74).

#### 4. Kesimpulan

Berdasarkan hasil dan pembahasan, berikut adalah simpulan dari metode yang digunakan. Setelah merancang dan menguji sistem analisis sentimen untuk ulasan film menggunakan metode Naive Bayes, diperoleh beberapa kesimpulan. Hasil klasifikasi menunjukkan bahwa model memiliki kinerja yang baik dalam mendeteksi sentimen positif, dengan *recall* yang sangat tinggi (0.99) dan akurasi keseluruhan sebesar 81%. Namun, model kurang efektif dalam mendeteksi sentimen negatif dan sepenuhnya gagal mengklasifikasikan sentimen netral. Ketidakeimbangan kelas dalam *dataset*, di mana sentimen positif mendominasi, kemungkinan menjadi penyebab utama kinerja yang tidak merata di antara kelas-kelas tersebut. Hal ini terlihat dari nilai *F1-score* yang tinggi untuk kelas positif (0.89), tetapi sangat rendah untuk kelas negatif (0.26) dan netral (0.00). Berdasarkan hasil ini, dapat disimpulkan bahwa metode Naive Bayes memiliki kinerja yang baik dalam mendeteksi sentimen positif, namun kinerjanya kurang memadai dalam mendeteksi sentimen negatif dan netral. Dengan *recall* yang tinggi untuk sentimen positif, metode ini cocok untuk kasus-kasus di mana dominasi sentimen positif dapat memberikan hasil yang berarti. Namun, untuk mencapai performa yang lebih seimbang dalam klasifikasi berbagai sentimen, perlu dilakukan penyesuaian pada *dataset* agar hasil analisis sentimen menjadi lebih optimal.

#### Referensi

- [1] N. S. Fathullah, Y. A. Sari, P. P. Adikara, "Analisis Sentimen Terhadap Rating dan Ulasan Film dengan menggunakan Metode Klasifikasi Naïve Bayes dengan Fitur Lexicon-Based," 2020. [Online]. Available: <http://j-ptiik.ub.ac.id>. [Accessed: Aug. 27, 2024].
- [2] M. H. Rifqo, A. Wijaya, "Implementasi Algoritma Naive Bayes dalam Penentuan Pemberian Kredit," 2017. [Online]. Available: [www.ejournal.unib.ac.id/index.php/pseudocode](http://www.ejournal.unib.ac.id/index.php/pseudocode). [Accessed: Aug. 27, 2024].
- [3] R. I. Pristiyanti, M. A. Fauzi, L. Muflikhah, "Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>. [Accessed: Aug. 27, 2024].
- [4] F. V. Sari, A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol. 10, no. 2, 2019. [Accessed: Aug. 27, 2024].
- [5] Karlinah, Jaja, "Analisis Teks Ulasan Film dari Media Elektronik Google dan Pemanfaatannya sebagai Bahan Ajar untuk Siswa SMP/MTS Kelas VIII," *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, vol. 1, 2019. [Accessed: Aug. 27, 2024].
- [6] I. Subagyo, L. Dwi Yulianto, W. Permadi, A. W. Dewantara, A. D. Hartanto, "Sentiment Analisis Review Film Di IMDB Menggunakan Algoritma SVM Sentiment Analysis of Film Review at IMDB using SVM algorithm," 2019. [Accessed: Aug. 27, 2024].
- [7] Emianti Hasibuan and Elmo Allistair Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier", *JTS*, vol. 1, no. 3, pp. 13–24, Oct. 2022. [Accessed: Aug. 27, 2024].
- [8] R. Wahyudi et al., "Analisis Sentimen pada Review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *Jurnal Informatika*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejournal/index.php/ji>. [Accessed: Aug. 27, 2024].
- [9] Muhammad Thaaqiq Razaq, D. Nurjanah, H. Nurrahmi, "Analisis Sentimen Review Film Menggunakan Naive Bayes Classifier Dengan Fitur TF-IDF," vol. 10, 2023. [Accessed: Aug. 27, 2024].
- [10] E. Fitri, Y. Yuliani, S. Rosyida, W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest dan Support Vector Machine," *TRANSFORMTIKA*, vol. 18, no. 1, pp. 71–80, 2020, [Online]. Available: [www.nusamandiri.ac.id](http://www.nusamandiri.ac.id). [Accessed: Aug. 27, 2024].