

---

## Perbandingan Kinerja Algoritma *Multinomial Naïve Bayes* dan *Logistic Regression* pada Analisis Sentimen Movie Ratings IMDB

Zulfah Binti Toyibah<sup>1</sup>, Yiyin Noriyah Putri<sup>2</sup>, Puandini<sup>3</sup>, Zalsa Maulina Widodo<sup>4</sup>, Ana Tsalitsatun Ni'mah<sup>5\*</sup>

<sup>1,2,3,4,5</sup> Program Studi Pendidikan Informatika, Universitas Trunojoyo Madura, Indonesia  
email: <sup>5</sup>[ana.tsalits@trunojoyo.ac.id](mailto:ana.tsalits@trunojoyo.ac.id)

DOI: <https://doi.org/10.21107/edutic.v10i2.28150>

---

### Abstrak

Salah satu hal yang dapat mempengaruhi seseorang memutuskan untuk menonton sebuah film adalah rating dari film itu sendiri. IMDB merupakan sebuah basis data daring yang berisikan informasi yang berkaitan dengan film, acara televisi, video rumahan, dan permainan video, dan acara internet, termasuk daftar pemeran, biografi kru produksi dan personil, ringkasan alur cerita, trivia, dan ulasan serta penilaian oleh penggemar. Ulasan yang diberikan oleh penggemar dapat berupa ulasan yang bersifat positif maupun negatif dari film yang telah ditonton. Penelitian ini bertujuan untuk mengetahui perbandingan akurasi dari algoritma *Multinomial Naïve Bayes* dan *Logistic Regression* dengan melakukan analisis sentimen pada data ulasan film oleh penggemar. Hasil dari pengujian komparasi ditemukan bahwa algoritma *Logistic Regression* memiliki kinerja yang terbaik dengan nilai akurasi 89.32%, sedangkan algoritma *Multinomial Naïve Bayes* memiliki nilai akurasi 85.28%. Sehingga dapat disimpulkan bahwa algoritma *Logistic Regression* memiliki nilai yang lebih baik dibandingkan dengan algoritma *Multinomial Naïve Bayes*.

**Kata Kunci:** Analisis Sentimen, *Logistic Regression*, *Multinomial Naïve Bayes*

### Abstract

One of the things that can influence a person's decision to watch a film is the rating of the film itself. IMDB is an online database containing information relating to movies, television shows, home videos, video games, and internet shows, including cast lists, biographies of the production crew and personnel, storyline summaries, trivia, and fan reviews and ratings. Reviews given by fans can be in the form of positive or negative reviews of the films that have been watched. This study aims to compare the accuracy of the *Multinomial Naïve Bayes* and *Logistic Regression* algorithms by conducting sentiment analysis on film review data by fans. The results of the comparative test found that the *Logistic Regression* algorithm has the best performance with an accuracy value of 89.32%, while the *Multinomial Naïve Bayes* algorithm has an accuracy value of 85.28%. So it can be concluded that the *Logistic Regression* algorithm has a better value than the *Multinomial Naïve Bayes* algorithm.

**Keywords:** Analisis Sentimen, *Logistic Regression*, *Multinomial Naïve Bayes*



© Author (s)

## PENDAHULUAN

Era digital saat ini, analisis sentimen telah menjadi topik yang sangat menarik dalam bidang pengolahan bahasa dan pemrosesan data. Pada analisis sentimen, tujuan utamanya adalah mengklasifikasikan suatu teks atau kalimat ke dalam kategori dan sesuai dengan sentimen yang terdapat di dalamnya, seperti positif dan negatif. Dalam konteks ini, algoritma klasifikasi seperti *Multinomial Naive Bayes* (MNB) dan *Logistic Regression* (LR) digunakan dalam melakukan analisis sentimen pada dataset rating film IMDB. Dalam analisis sentimen rating film IMDB, menggunakan dua algoritma populer yang digunakan, yaitu *Multinomial Naive Bayes* (MNB) dan *Logistic Regression* (LR). Kedua algoritma ini digunakan untuk mengklasifikasikan ulasan film sebagai positif atau negatif berdasarkan sentimen yang terkandung dalam teks tersebut berdasarkan ulasan pengguna terhadap film-film di IMDB.

*Multinomial Naive Bayes* adalah algoritma dengan satu metode pembelajaran probabilistik yang berdasarkan teorema bayes dan biasanya digunakan pada *Natural Language Processing* (NLP) (Yuyun et al., 2021). Algoritma ini telah banyak digunakan pada pemrosesan data teks dan analisis sentimen karena efisiensinya dan kemampuannya dalam mengatasi masalah dimensi tinggi pada data teks. Di sisi lain, algoritma *Logistic Regression* adalah model statistik yang digunakan untuk memprediksi probabilitas biner (dua kelas) atau multi-klas dari suatu peristiwa berdasarkan fitur-fitur yang ada. Dalam konteks analisis sentimen, *Logistic Regression* dapat digunakan untuk memprediksi sentimen positif atau negatif berdasarkan fitur-fitur yang terdapat dalam teks.

Perbandingan kinerja antara MNB dan LR pada analisis sentimen rating film IMDB dapat dipengaruhi oleh beberapa faktor. Pertama, jumlah dan kualitas data pelatihan yang digunakan dapat mempengaruhi kinerja kedua algoritma. Semakin banyak dan berkualitas data pelatihan yang digunakan, semakin baik kinerja algoritma dalam mempelajari pola sentimen yang ada. Perbandingan kinerja antara MNB dan LR pada analisis sentimen rating film IMDB dapat dilakukan melalui evaluasi menggunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score. Dengan melakukan evaluasi yang cermat, dapat diketahui algoritma mana yang lebih baik dalam mengklasifikasikan sentimen pada rating film IMDB. Perbandingan kinerja antara algoritma *Multinomial Naive Bayes* dan *Logistic Regression* dalam analisis sentimen rating film IMDB dapat memberikan wawasan tentang kelebihan dan kelemahan masing-masing algoritma dalam menangani tugas tersebut. Beberapa faktor yang dapat diperhatikan dalam perbandingan ini antara lain akurasi klasifikasi, interpretabilitas model, dan kestabilan kinerja algoritma terhadap perubahan dataset. Selanjutnya, pemilihan fitur-fitur yang relevan juga dapat mempengaruhi kinerja algoritma. Fitur-fitur ini dapat berupa kata-kata, frasa, atau fitur lain yang digunakan untuk mewakili teks ulasan. Penggunaan fitur-fitur yang informatif dan relevan dapat meningkatkan kinerja algoritma.

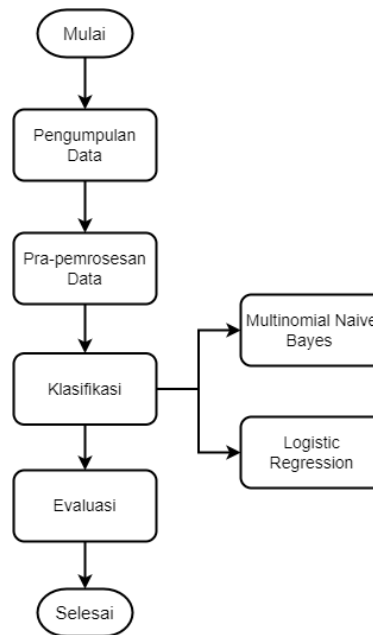
Jumlah dan kualitas data yang cukup dan representatif sangat penting untuk meningkatkan kinerja algoritma. Semakin banyak data yang tersedia, semakin baik algoritma dapat mempelajari pola-pola sentimen yang ada dalam data. Selain itu, pemilihan fitur-fitur yang relevan juga dapat mempengaruhi kinerja. Fitur-fitur ini dapat berupa kata-kata kunci, frasa, atau fitur lain yang dapat mengungkapkan sentimen secara efektif. Pengaturan parameter yang tepat juga memainkan peran penting dalam perbandingan kinerja. Sebagai contoh, sebuah jurnal yang diterbitkan pada tahun 2018 oleh Ruchi and Rahul Sharma dengan judul "*Comparative Study of Sentiment Analysis using Naive Bayes, SVM and Logistic Regression*" melakukan perbandingan antara algoritma *Multinomial Naive Bayes*, *Support Vector Machine* (SVM), dan *Logistic Regression* pada analisis sentimen menggunakan dataset IMDB movie reviews. Hasil dari studi ini menunjukkan bahwa SVM memberikan akurasi tertinggi, diikuti oleh Logistic Regression, dan kemudian Multinomial Naive Bayes. Selanjutnya penelitian yang dilakukan oleh (Perdana et al., n.d.) dengan judul "Komparasi Algoritma *Naive Bayes*, *Support Vector Machine*, Dan *Logistic Regression* pada Analisis Sentimen Pengguna Aplikasi

Transportasi Online”. Dari hasil penelitian diperoleh akurasi tertinggi sebesar 88.5% dengan algoritma *Naïve Bayes*.

Penelitian terdahulu inilah yang menjadikan kami untuk meneliti lebih lanjut dengan dataset dan algoritma yang berbeda. Dengan tujuan membuktikan bahwa algoritma tersebut bisa dilatih lebih lanjut sehingga menghasilkan akurasi yang tinggi. Dengan perbandingan kedua algoritma ini bertujuan agar dapat mengetahui keunggulan dan kelemahan dari masing- masing model. Namun, perlu digaris bawahi bahwa hasil perbandingan kinerja ini dapat berbeda-beda tergantung pada berbagai faktor seperti ukuran dataset, pemrosesan teks, pemilihan fitur, pengaturan parameter, dan metrik evaluasi yang digunakan.

**METODE PENELITIAN**

Metodologi penelitian merupakan gagasan kerangka penelitian yang disusun secara sistematis sehingga saat melakukan sebuah penelitian, membahas masalah yang muncul pada rumusan pertanyaan penelitian serta mencapai hasil yang diinginkan. Metode penelitian yang dilakukan terdiri dari beberapa tahapan yaitu pengumpulan data, pra-pemrosesan data, klasifikasi dengan menggunakan perbandingan dua algoritma yaitu *Multinomial Naïve Bayes* dan *Logistic Regression* serta tahap akhir yaitu evaluasi. Untuk lebih jelasnya dapat dilihat pada Gambar 1 berikut.



**Gambar 1.** Metodologi Alur Penelitian

**2.1 Pengumpulan Data**

Proses pengumpulan data dalam penelitian ini mengambil dataset dari *Kaggle*. Data yang diambil sejumlah 40.000 yang diambil dari rating pengguna terhadap movie IMDB. Kemudian data itu disimpan dalam file bertipe .csv agar data bisa digunakan dan dapat terbaca secara baik pada read data csv, data set ini terdiri dari dua atribut yaitu text dan label.

**2.2 Pra-pemrosesan data**

Tahap pre-processing data, data mentah pertama-tama melalui proses *case folding*, *tokenizing*, *stemming*, *stopwords removal*, serta TF-IDF. Tahap Pre-processing ini dilakukan untuk menghindari data yang tidak teratur, tidak lengkap dan tidak konsisten sehingga dalam proses selanjutnya yaitu klasifikasi model bisa berjalan dengan baik.

**2.3 Klasifikasi**

Tahapan klasifikasi merupakan bagian tahap yang utama pada proses analisis sentimen yang termasuk kedalam tipe *supervised learning*. Istilah klasifikasi didapat dari tujuan utama dalam teknik ini yaitu untuk memprediksi kategori dari masukkan data (Supriyatna & Mustika, 2018). Tujuan dari proses klasifikasi ini untuk perhitungan dengan menerapkan algoritma data mining dan mengolah dataset ulasan dari ratings movie IMDB. Terdapat beberapa algoritma yang digunakan antara lain *Multinomial Naïve Bayes* dan *Logistic Regression*.

### 2.3.1 *Multinomial Naïve Bayes*

Algoritma *Naïve Bayes* adalah salah satu algoritma dalam teknik klasifikasi, algoritma ini memanfaatkan metode probabilitas dan statistik (Perdana et al., n.d.). Algoritma *Naïve Bayes* disebut sebagai algoritma pembelajaran sederhana dengan menggunakan aturan bayes dalam mengkalsifikasikan atau memprediksi suatu probabilitas class data (Yuda Lesmana & Andarsyah, 2022). Algoritma ini merupakan salah satu algoritma yang sangat populer dan termasuk dalam sepuluh algoritma teratas pada data mining menurut IEEE (International Conference on Data Mining) di Hongkong (Giovani et al., 2020). Salah satu tipe dari algoritma *Naïve Bayes* adalah *Multinomial Naïve Bayes*. Pengklasifikasian algoritma *Multinomial Naïve Bayes* ini sangat cocok digunakan dalam klasifikasi fitur diskrit (seperti, jumlah kata dalam klasifikasi teks). Berikut adalah persamaan dari algoritma *Multinomial Naïve Bayes*.

$$P(H|X) = \frac{P(H) * P(X|H)}{P(X)}$$

Keterangan:

- X = Data kelas tidak diketahui
- H = Data hipotesis X adalah kelas yang terpisah
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi (posteriori probability)
- P(H) = Probabilitas hipotesis H
- P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis
- P(X) = Probabilitas X

### 2.3.2 *Logistic Regression*

Algoritma *Logistic Regression* merupakan suatu jenis metode analisis yang digunakan untuk mendeskripsikan hubungan antara variabel terikat yang memiliki dua kategori atau lebih dengan satu atau lebih variabel bebas, dengan skala kategori atau kontinu (Lengkong et al., 2021). Algoritma ini digunakan untuk menjelaskan hubungan antara variabel independen dan variabel dependen dengan menghubungkan satu ataupun lebih variabel bebas dan variabel terikat yang jenis kategorinya bisa berupa 0 dan 1, true atau false, besar atau kecil. Tipe variabel bebasnya berupa kategori sehingga inilah yang membedakan algoritma ini dari regresi berganda ataupun regresi linier lainnya (Pramakrisna et al., 2022). Berikut adalah persamaan dari algoritma *Logistic Regression*.

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_1X$$

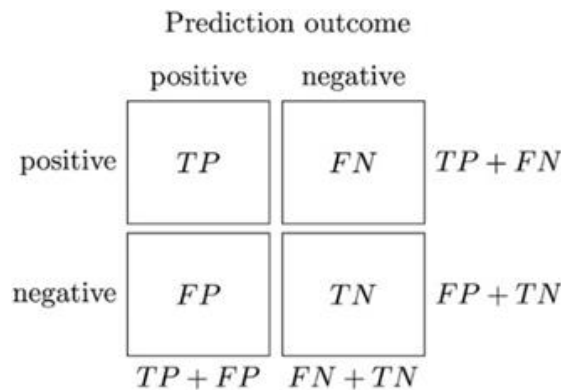
Keterangan:

- B0 = Konstanta
- B1 = Koefisien dari masing-masing variabel
- Nilai p atau peluang (Y = 1) dapat dicari dengan persamaan

$$p = \frac{e(B_0 + B_1X)}{(1 + e(B_0 + B_1X))}$$

## 2.4 Evaluasi

Tahapan ini dilakukan untuk membuktikan bahwa tes yang dilakukan benar. Evaluasi digunakan untuk menemukan hasil yang dinilai terbaik dari hasil uji. Kriteria evaluasi yang dipertimbangkan adalah akurasi, recall, precision, standard deviation, f1 score, dan specificity (Attal et al., 2015). Pengukuran akurasi terhadap model menggunakan Confusion Matrix. Confusion Matrix merupakan suatu alat yang digunakan untuk menganalisis seberapa model klasifikasi mengenali tuple dari suatu data yang berbeda (Nasution & Hayaty, 2019).



**Gambar 2.** Confusion Matrix

Keterangan:

TP (*True Positive*) = Adalah data positif diklasifikasikan secara benar

TN (*True Negative*) = Adalah data negatif diklasifikasikan secara benar

FP (*False Positive*) = Adalah data negatif diklasifikasikan secara positif

FN (*False Negative*) = Adalah data positif diklasifikasikan secara negatif

## HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Data

Proses yang dilakukan pertama kali adalah pengambilan data pada ulasan rating film platform IMDB. Data yang dipakai pada penelitian ini merupakan data yang sudah di ditambah dan sudah diberi label menjadi dua kelas yakni 0 dan 1. Kelas 0 digunakan untuk mewakili label negatif dan kelas 1 digunakan untuk mewakili label positif. Data ini di dapat dari platform Kaggle. Dataset yang digunakan dalam penelitian ini sebanyak 40.000 data. Dataset dapat dilihat pada gambar 3:

id	text	label
1		
2	I grew up (b. 1965) watching and loving the Thunderbirds. All my	0
3	When I put this movie in my DVD player, and sat down with a co	0
4	Why do people who do not know what a particular time in the p	0
5	Even though I have great interest in Biblical movies, I was bored	0
6	Im a die hard Dads Army fan and nothing will ever change that. I	1
7	A terrible movie as everyone has said. What made me laugh wa:	0
8	Finally watched this shocking movie last night, and what a distu	1
9	I caught this film on AZN on cable. It sounded like it would be a j	0
10	It may be the remake of 1987 Autumn's Tale after eleven years,	1
11	My Super Ex Girlfriend turned out to be a pleasant surprise for n	1
12	I can't believe people are looking for a plot in this film. This is La	1
13	If you haven't seen the gong show TV series then you won't like	0
14	I have always been a huge fan of "Homicide: Life On The Street"	1
15	Greg Davis and Bryan Daly take some crazed statements by a ter	0
16	A half-hearted attempt to bring Elvis Presley into the modern di	0
17	If you want a fun romp with loads of subtle humor, then you wil	1

**Gambar 3.** Dataset

Penelitian ini menggunakan perangkat keras dan perangkat lunak dengan spesifikasi sebagai berikut.

**Tabel 1.** Spesifikasi perangkat yang digunakan

Hardware	Spesifikasi
----------	-------------

Processor	Intel® Core™ i7-8550U CPU @ 1.80GHz (8 CPUs), ~2.0GHz
Tipe Sistem	64-bit
Operation System	Windows 10 Home
Harddisk	1 TB
Memori (RAM)	8 GB
Software	<ul style="list-style-type: none"> <li>• Google Colab</li> <li>• Google Drive</li> </ul>

### 3.2. Pra-Pemrosesan Data

Setelah melakukan tahap pengumpulan data, proses selanjutnya adalah melakukan cleaning data atau pembersihan data dengan tahapan pra-pemrosesan dalam penelitian ini. Tahapan pra-pemrosesan ini berisi tahapan *case folding*, *stopword removal*, *stemming*, TF-IDF, dan pembagian dataset dengan rasio 80:20. Dalam tahapan ini untuk tahapan *case folding* dijadikan satu menjadi function seperti berikut.

```
[8] def preprocessor(text):
    """ Return a cleaned version of text
    """
    # Remove HTML markup
    text = re.sub('<[^>*>', '', text)
    # Save emoticons for later appending
    emoticons = re.findall('(?:;|=)(?:-)?(?:\)|\(|D|P)', text)
    # Remove any non-word character and append the emoticons,
    # removing the nose character for standarization. Convert to lower case
    text = (re.sub('[\W]+', ' ', text.lower()) + ' ' + ' '.join(emoticons).replace('-', ''))

    return text
```

Gambar 4. Source Code tahapan *case folding*

Tahapan ini adalah penghapusan karakter dari HTML Markup, penghapusan emoticon, penghapusan angka, penghapusan karakter, dan mengubah data menjadi huruf kecil semua. Untuk tahapan *stemming* juga dijadikan menjadi satu function.

```
[9] from nltk.stem import PorterStemmer

porter = PorterStemmer()

def tokenizer_porter(text):
    token = []
    for word in text.split():
        token.append(porter.stem(word))

    return token
```

Gambar 5. Source Code tahapan *stemming*

Tahap ini terdapat penghilangan semua kata imbuhan yang ada pada kata tersebut dan diubah menjadi kata dasar. Untuk tahapan *stopword removal* juga dijadikan menjadi satu function. Selain itu Pada tahapan ini selain terdapat *stopword removal*, terdapat juga *tokenizing* dan pembobotan TF-IDF.

```
[10] from sklearn.feature_extraction import text
      from sklearn.feature_extraction.text import TfidfVectorizer
      from nltk.corpus import stopwords
      import nltk
      nltk.download('stopwords')

      from sklearn.model_selection import train_test_split

      X = movies['text']
      y = movies['label']

      my_additional_stop_words = []

      stop = text.ENGLISH_STOP_WORDS.union(my_additional_stop_words)

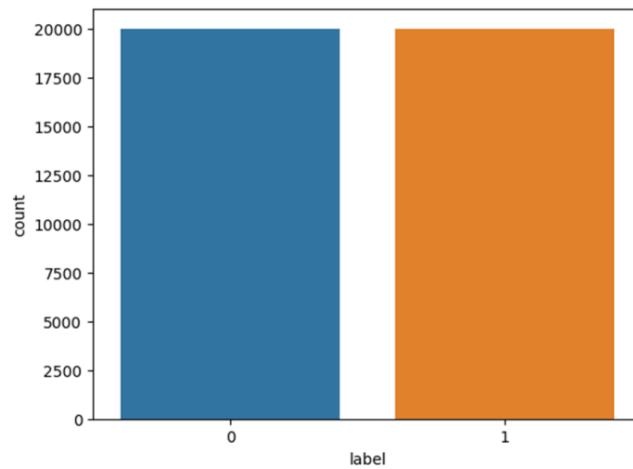
      tfidf = TfidfVectorizer(stop_words=list(stop),
                             tokenizer=tokenizer_porter,
                             preprocessor=preprocessor)

      X_tfidf = tfidf.fit_transform(X)
```

Gambar 6. Source Code tahapan *stopword removal*, *tokenizing*, dan *TF-IDF*

### 3.3. Klasifikasi

Setelah selesai tahapan pra-pemrosesan maka selanjutnya melakukan proses klasifikasi. Proses ini menggunakan perangkat lunak yaitu *Google Colab* dengan Bahasa Pemrograman yang digunakan *Python* . Pada proses klasifikasi ini data sebanyak 40.000 data dibagi menjadi 2 kelas yaitu 0 dan 1. Hasil perbandingan kelasnya dapat dilihat pada gambar di bawah ini.



Gambar 7. Visualisasi data hasil klasifikasi sentimen Positif dan Negatif

Hasil klasifikasi data yang telah dilakukan didapatkan data dengan kelas 0 (negatif) sebanyak 20.019 data dan data dengan kelas 1 (positif) sebanyak 19.981 sehingga dari kedua kelas ini data tergolong seimbang. Setelah itu data dibagi menjadi 2 yaitu data training dan data testing dengan perbandingan ratio 0.2. Artinya data training berjumlah 80% yaitu sebanyak 32.000 data dan data testing berjumlah 20% yaitu sebanyak 8.000 data. Hasil akurasi klasifikasi data testing yang berisi analisis sentiment rating film di platform *IMDB* menggunakan *Multinomial Naïve Bayes* dan *Logistic Regression* disajikan pada tabel di bawah ini.

Tabel 2. Hasil perbandingan performa algoritma

NO	Algoritma	Akurasi(%)	Precision(%)	Recall(%)
1	<i>Multinomial Naïve Bayes</i>	85.28	86.84	83.05
2	<i>Logistic Regression</i>	89.32	88.15	90.80

Berdasarkan tabel di atas algoritma pertama yaitu *Multinomial Naïve Bayes* mendapatkan akurasi sebanyak 85.28%, Precision 86.84%, Recall 83.05%. Sedangkan algoritma kedua *Logistic Regression* mendapatkan akurasi 89.32%, Precision 88.15%, Recall 90.80%. Dapat disimpulkan bahwa nilai performa akurasi model algoritma *Logistic Regression* mendapatkan nilai akurasi lebih baik yaitu 89.32% dengan selisih 4.04% dibandingkan dengan algoritma *Multinomial Naïve Bayes*.

### 3.4. Evaluasi

Tahap terakhir ini dimana menghitung ulang hasil dari nilai akurasi yang didapat oleh masing-masing algoritma menggunakan *Confusion Matrix*.

**Tabel 3.** *Confusion Matrix Logistic Regression*

	Positif	Negatif
Positif	3525	487
Negatif	367	3621

Berdasarkan tabel di atas dapat dilihat bahwa jumlah TP sebanyak 3525, FP sebanyak 367, FN sebanyak 487, dan TN sebanyak 3621. Cara menghitung manualnya adalah dengan  $TP+FP+FN+TN = 8.000$ , Berikutnya  $(TP+TN)/8000 = 0.8932$ . Hasil dari perhitungan confusion matrix logistic regression adalah 89.32%.

**Tabel 4.** *Confusion Matrix Multinomial Naïve Bayes*

	Positif	Negatif
Positif	3510	502
Negatif	676	3312

Berdasarkan tabel di atas dapat dilihat bahwa jumlah TP sebanyak 3510, FP sebanyak 676, FN sebanyak 502, dan TN sebanyak 3312. Cara menghitung manualnya adalah dengan  $TP+FP+FN+TN = 8.000$ , Berikutnya  $(TP+TN)/8000 = 0.8527$ . Hasil dari perhitungan confusion matrix Multinomial Naïve Bayes adalah 85.27%.

## KESIMPULAN

Hasil dari pengujian yang telah dilakukan ditunjukkan penggunaan algoritma *Multinomial Naïve Bayes* dan *Logistic Regression* dapat digunakan pada klasifikasi analisis sentimen movie ratings IMDB. Berdasarkan hasil uji komparasi didapatkan bahwa algoritma *Logistic Regression* memiliki kinerja lebih unggul dibandingkan algoritma *Multinomial Naïve Bayes*. Dibuktikan dengan hasil akurasi yang didapatkan pada algoritma *Logistic Regression* sebesar 89.32% sedangkan algoritma *Multinomial Naïve Bayes* memiliki akurasi jauh lebih rendah yaitu 85.28%. Jumlah dataset yang digunakan sebanyak 40.000 data yang tersedia pada Kaggle, kemudian data dibagi menjadi 2 yaitu data testing sebanyak 8000 data dan data training sebanyak 32.000 data dengan rasio perbandingan 20:80. Sehingga diperoleh kesimpulan bahwa algoritma *Logistic Regression* memiliki kinerja yang lebih baik dibandingkan algoritma *Multinomial Naïve Bayes*.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ibu Ana Tsalitsatun Ni'mah, S.Kom., M.Kom. selaku dosen pengampu mata kuliah *Text Mining* kelas Program Studi Pendidikan Informatika, Fakultas Ilmu Pendidikan, Universitas Trunojoyo Madura.

## DAFTAR PUSTAKA

Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y. (2015).



- 
- Physical human activity recognition using wearable sensors. *Sensors (Switzerland)*, 15(12), 31314–31338. <https://doi.org/10.3390/s151229858>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Lengkong, N. C., Safitri, O., Machsus, S., Putra, Y. R., Syahadati, A., & Nooraeni, R. (2021). Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg. *Jurnal Teknoinfo*, 15(1), 20. <https://doi.org/10.33365/jti.v15i1.866>
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 226–235. <https://doi.org/10.31311/ji.v6i2.5129>
- Perdana, K., Sitompul, J., Pratama, A. R., & Baihaqi, K. A. (n.d.). *VECTOR MACHINE , DAN LOGISTIC REGRESSION PADA ANALISIS SENTIMEN PENGGUNA APLIKASI*. 27–38.
- Pramakrisna, F. D., Adhinata, F. D., & Tanjung, N. A. F. (2022). Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression. *Teknika*, 11(2), 90–97. <https://doi.org/10.34148/teknika.v11i2.466>
- Supriyatna, A., & Mustika, W. P. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kulit. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 2(2), 152. <https://doi.org/10.30645/j-sakti.v2i2.78>
- Yuda Lesmana, R., & Andarsyah, R. (2022). Model Klasifikasi Multinomial Naïve Bayes Untuk Analisis Sentiment Terkait Non-Fungible Token. *Jurnal Teknik Informatika*, 14(3), 135–139.
- Yuyun, Nurul Hidayah, & Supriadi Sahibu. (2021). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 820–826. <https://doi.org/10.29207/resti.v5i4.3146>