

IMPLEMENTASI *NAIVE BAYES CLASSIFIER* (NBC) UNTUK MENGANALISIS SENTIMEN ULASAN FILM “*EVERYTHING EVERYWHERE ALL AT ONCE*” DI TWITTER

ALDI FLORENSIUS GEA¹, CHOIRUL BASIR^{2*}

^{1,2}Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pamulang

*dosen02278@unpam.ac.id

ABSTRAK

Media sosial, khususnya Twitter, telah menjadi platform penting untuk berbagi informasi dan pandangan, termasuk di Indonesia. Penelitian ini berfokus pada analisis sentimen ulasan film “*Everything Everywhere All at Once*” di Twitter menggunakan metode *Naïve Bayes Classifier* (NBC). Tujuan penelitian ini adalah untuk mendapatkan nilai akurasi klasifikasi berbentuk *confusion matrix* pada ulasan film “*Everything Everywhere All at Once*” di Twitter menggunakan metode *Naïve Bayes Classifier* (NBC), mengetahui hasil penerapan metode NBC dalam mengklasifikasikan data ulasan film menjadi label kelas positif dan negatif, serta visualisasi persepsi positif dan negatif pada film tersebut. Data yang digunakan terdiri dari 2148 tweet yang dikumpulkan antara 1 Juli 2022 hingga 16 Desember 2023. Hasil penelitian menunjukkan bahwa model NBC berhasil mencapai akurasi 73%, dengan 1388 ulasan positif dan 760 ulasan negatif. Selain itu, hasil penerapan metode NBC secara manual juga menunjukkan akurasi yang sama. Visualisasi dalam bentuk wordcloud mengungkap kata-kata yang sering muncul, seperti “*everything*,” “*everywhere*,” “*bagus*,” dan “*recommended*” untuk ulasan positif, serta “*ngantuk*,” “*tidur*,” dan “*bingung*” untuk ulasan negatif. Penelitian ini diharapkan dapat memberikan wawasan kepada pembuat film, peneliti, dan pengembang teknologi dalam memahami persepsi publik terhadap film melalui analisis media sosial.

Kata Kunci: Analisis sentimen, *Confusion Matrix*, Film, *Naïve Bayes Classifier*, Twitter

ABSTRACT

Social media, especially Twitter, has become an important platform for sharing information and views, including in Indonesia. This research focuses on sentiment analysis of “Everything Everywhere All at Once” movie reviews on Twitter using the Naïve Bayes Classifier (NBC) method. The purpose of this research is to obtain the classification accuracy value in the form of a confusion matrix on the “Everything Everywhere All at Once” movie review on Twitter using the Naïve Bayes Classifier (NBC) method, knowing the results of the application of the NBC method in classifying movie review data into positive and negative class labels, and visualizing positive and negative perceptions of the film. The data used consists of 2148 tweets collected between July 1, 2022 and December 16, 2023. The results showed that the NBC model successfully achieved 73% accuracy, with 1388 positive reviews and 760 negative reviews. In addition, the results of manually applying the NBC method also show the same accuracy. Visualization in the form of a wordcloud reveals frequently occurring words, such as “everything,” “everywhere,” “good,” and “recommended” for positive reviews, as well as “sleepy,” “sleeping,” and “confused” for negative reviews. This research is expected to provide insights to filmmakers, researchers, and technology developers in understanding public perception of movies through social media analysis.

Keywords: *Confusion Matrix, Movies, Naïve Bayes Classifier, Sentiment analysis, Twitter*

1 Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah memberikan kontribusi signifikan terhadap penyediaan layanan web, termasuk media sosial [1]. Media sosial seperti Twitter, Facebook, dan YouTube menjadi platform utama untuk berkomunikasi, menyebarkan informasi, dan mengekspresikan diri [2]. Pertumbuhan pengguna media sosial semakin pesat, tidak hanya dari segi jumlah pengguna tetapi juga dalam hal kegunaannya [3]. Media sosial terus berkembang menjadi sarana penting dalam kehidupan sehari-hari masyarakat global.

Twitter, yang sekarang dikenal sebagai aplikasi X, merupakan salah satu media sosial yang paling sering digunakan untuk berbagi informasi secara *real-time*. Menurut penelitian SemioCast pada tahun 2013, Indonesia merupakan negara dengan pengguna Twitter terbanyak ketiga di dunia, setelah Amerika Serikat dan Jepang, dengan 6,5 persen pengguna global dan satu miliar tweet [4]. Dengan peningkatan jumlah pengguna, volume tweet yang diposting juga meningkat signifikan. Pengguna Twitter sering membahas berbagai topik, termasuk menonton film, dan mengungkapkan pendapat mereka tentang film tersebut secara langsung [5].

Film merupakan media komunikasi audio-visual yang memiliki kemampuan untuk menyampaikan pesan tersirat kepada penonton. Pesan yang disampaikan melalui film dapat mempengaruhi persepsi penonton dan mendorong mereka untuk mengungkapkan opini mereka mengenai alur, latar, penokohan, dan akhir cerita. Dalam dunia perfilman, ada beberapa ajang penghargaan bergengsi seperti *Academy Awards* (Piala Oscar), *BAFTA Awards*, dan *Golden Globe Awards* yang mengapresiasi pencapaian luar biasa dalam industri film. Penghargaan ini menjadi tolak ukur kesuksesan sebuah film di mata publik.

Pada tahun 2023, film "*Everything Everywhere All at Once*" menjadi salah satu film yang paling banyak diperbincangkan setelah berhasil memenangkan tujuh penghargaan di ajang Piala Oscar, termasuk kategori film terbaik. Kesuksesan ini menarik perhatian banyak penonton, yang kemudian memberikan berbagai macam ulasan mengenai film tersebut. Berdasarkan data dari IMDb, film ini mendapat rating 7,8/10 dari lebih dari 542.000 pengguna, menunjukkan bahwa film ini memiliki daya tarik yang signifikan di kalangan penonton. Berbagai pandangan yang berbeda mengenai film ini muncul, mencerminkan persepsi yang beragam dari masyarakat [6].

IMDb adalah salah satu situs web terbesar yang menyediakan informasi komprehensif tentang film, termasuk rating dan ulasan dari pengguna. Situs ini memengaruhi keputusan penonton dalam memilih film yang akan ditonton, karena mereka dapat melihat rating dan membaca ulasan sebelum memutuskan untuk menonton [7]. Di Indonesia, IMDb memiliki pangsa pengguna yang cukup besar, dengan 1,3 persen dari 57 juta pengunjung global. Studi menunjukkan bahwa rating di IMDb memiliki pengaruh signifikan terhadap keputusan seseorang untuk menonton film tertentu, sehingga menjadikannya sumber informasi yang sangat berharga bagi penonton.

Melihat *rating* dan popularitas film "*Everything Everywhere All at Once*" di IMDb, analisis sentimen diperlukan untuk memahami bagaimana pandangan masyarakat terhadap film ini. Analisis sentimen membantu mengidentifikasi apakah ulasan publik di media sosial cenderung positif atau negatif. Metode *Naïve Bayes* adalah salah satu teknik *machine learning* yang populer untuk klasifikasi teks dan dapat digunakan untuk mengolah dataset ulasan film ini. Dengan menggunakan *Naïve Bayes*, komentar di media sosial dapat dikelompokkan menjadi kategori positif atau negatif, memberikan wawasan yang lebih mendalam tentang reaksi publik terhadap film tersebut [8].

2 Tinjauan Pustaka

2.1 Film

Film adalah suatu bentuk komunikasi massa elektronik yang berupa media audio visual yang mampu menampilkan kata-kata, bunyi, citra, dan kombinasinya. Film juga merupakan salah satu bentuk komunikasi modern yang kedua muncul di dunia. Film berperan sebagai sebuah sarana baru yang digunakan untuk menyebarkan hiburan yang sudah menjadi kebiasaan terdahulu, serta menyajikan cerita, peristiwa, musik, drama, lawak, dan sajian teknis lainnya kepada masyarakat umum. Film juga menurut [9] adalah media komunikasi massa yang ampuh sekali, bukan saja untuk hiburan, tetapi juga untuk penerangan dan pendidikan. Film mempunyai suatu dampak tertentu terhadap penonton, dampak – dampak tersebut dapat berbagai macam seperti, dampak psikologis, dan dampak sosial [10].

2.2 Twitter

Twitter adalah salah satu inovasi teknologi komunikasi daring yang fenomenal di abad 21 ini. Platform ini dikenal bukan hanya sebagai penyedia layanan media sosial, tetapi juga diidentifikasi sebagai bentuk microblogging. Dalam Jurnal Analisis Penggunaan Twitter Sebagai Media Komunikasi Selebritis di Jakarta, [11] menjelaskan bahwa Twitter hadir sebagai wujud perkembangan teknologi yang dimanfaatkan oleh manusia modern untuk berkomunikasi. Twitter menjadi salah satu dari banyak media sosial yang memikat perhatian pengguna internet karena kemudahan penggunaannya dalam pertukaran informasi, memungkinkan setiap individu di seluruh dunia untuk saling terhubung.

2.3 Sentimen Analisis

Sentimen analisis adalah proses otomatis untuk memahami, mengekstrak, dan mengolah data teks guna mengidentifikasi sentimen yang terkandung dalam sebuah kalimat opini. Metode ini digunakan untuk menilai pendapat atau kecenderungan opini seseorang terhadap suatu isu atau objek, apakah cenderung positif atau negatif.

2.4 Preprocessing Data

Text preprocessing adalah langkah penting dalam analisis sentimen, yang dapat memengaruhi akurasi algoritma *machine learning* [12]. Tujuannya adalah mengubah data mentah menjadi data terstruktur untuk analisis lebih lanjut. Langkah-langkah utama *text preprocessing* mencakup pembersihan, normalisasi, dan transformasi data, yang termasuk dalam tahap *modify* pada SEMMA. Secara khusus, *text preprocessing* meliputi *Spelling Normalization*, *case folding*, *word normalization*, *filtering*, *stemming*, dan *tokenizing* [13]. Proses ini mengubah teks yang tidak terstruktur menjadi terstruktur.

2.5 Naïve Bayes Classifier

Klasifikasi *Naïve Bayes* adalah teknik data mining yang populer untuk mengklasifikasi data dan memprediksi kelas. Teknik ini digunakan untuk membangun model yang menjelaskan dan membedakan kelas data, dengan tujuan memperkirakan kelas dari objek yang labelnya tidak diketahui. Klasifikasi *Naïve Bayes* mengasumsikan independensi yang kuat antar fitur dalam data, yang berarti keberadaan suatu fitur tidak mempengaruhi fitur lainnya dalam kelas yang sama. Karena kesederhanaannya, *Naïve Bayes* sering kali efektif dibandingkan dengan metode lain seperti *Decision Tree*, *Neural Network*, dan *Support Vector Machines (SVM)* dalam hal akurasi dan efisiensi komputasi. Metode ini menggunakan teorema Bayes dan dapat digunakan pada dataset besar serta sangat efektif dalam analisis sentimen.

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)} \quad (1)$$

dengan:

- X : Data dengan kelas yang belum diketahui
- Y : Hipotesis data X yang merupakan suatu kelas spesifik
- $P(Y|X)$: Probabilitas hipotesis Y berdasar kondisi X
- $P(Y)$: Probabilitas hipotesis Y
- $P(X|Y)$: Probabilitas X berdasarkan kondisi pada hipotesis Y
- $P(X)$: Probabilitas X

2.6 Evaluasi

Evaluasi adalah tahapan terakhir dari suatu proses analisis yaitu dimana membandingkan antara hasil implementasi dengan kriteria standar yang telah ditentukan. Selain itu evaluasi dapat mengetahui tingkat keberhasilan yang dicapai dan mengetahui selisih antara standar yang ditentukan dengan hasil yang dicapai. Evaluasi dapat dilakukan dengan menghitung akurasi, dilakukan hal ini dapat mengetahui kinerja sistem pada kasus klasifikasi [14,15]. Sistem dapat dihitung dengan tabel kontingensi. Berikut tabel kontingensi.

Tabel 1: Kontingensi

	<i>Classified positive</i>	<i>Classified negative</i>
<i>Actual positive</i>	TP	FN
<i>Actual negative</i>	FP	TN

Dengan:

- TP : *True Positive* yaitu model memprediksi benar bahwa itu kelas positif.
- TN : *True Negative* yaitu model memprediksi benar bahwa itu kelas negatif.
- FP : *False Positive* yaitu model memprediksi salah kelas positif tapi kenyataannya itu kelas negatif.
- FN : *False Negative* yaitu model memprediksi salah kelas negatif tapi kenyataannya itu kelas positif.

Ukuran evaluasi model yang terdiri dari akurasi yang dapat dihitung berdasarkan nilai dalam *confusion matrix* pada Tabel 2. untuk mengetahui tingkat akurasi yaitu sebagai berikut:

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2)$$

3 Hasil dan Pembahasan

3.1 Pengumpulan Data

Langkah awal penelitian ini adalah mengumpulkan data tweet ulasan film “*Everything Everywhere All at Once*” dalam bahasa Indonesia melalui *scraping* data dari Twitter. *Scraping* dilakukan menggunakan *library Tweet Harvest* di Python setelah mendapatkan *Auth Token* untuk mengakses *API* Twitter. Persiapan mencakup instalasi ‘*pandas*’ dan ‘*Node.js*’, serta mengimpor *library* yang dibutuhkan seperti ‘*Tweet Harvest*’, ‘*pandas*’, dan ‘*numpy*’. Proses pengumpulan data dilakukan dengan dua kali *scraping* menggunakan *query* “*Everything Everywhere All at Once*” dalam bahasa Indonesia, dengan rentang waktu 1 Juli 2022 - 3 Maret 2023, dan 3 Maret 2023 - 16 Desember 2023. Total 2148 data tweet yang relevan berhasil dikumpulkan dan disimpan dalam format CSV, setelah dilakukan *filtering* hanya menyisakan teks tweet saja.

3.2 Preprocessing Data

Preprocessing data merupakan tahapan yang harus dilakukan dalam proses penelitian ulasan ini, karena pada tahapan ini akan dilakukannya perubahan terhadap setiap kata pada ulasan yang disetiap kata pada ulasan tidak sepenuhnya menggunakan bahasa yang baku dan dengan melakukan *preprocessing* data ini akan merubahnya menjadi bahasa Indonesia yang lebih baik. Tahapan *preprocessing* data meliputi beberapa tahapan sehingga data menjadi bersih dan dapat digunakan untuk tahapan selanjutnya, berikut tahapan pada *preprocessing* data yaitu meliputi tahapan *Case folding*, *Tokenizing*, *Text Normalization*, *Filtering (stopword removal)*, dan *Stemming*.

Case folding adalah sebuah teknik dalam pemrosesan teks yang bertujuan untuk menyamakan penggunaan huruf besar dan kecil dalam sebuah teks. Teknik ini sangat penting dalam analisis teks karena membantu menghilangkan perbedaan yang disebabkan oleh variasi penggunaan huruf kapital dan kecil. Dalam konteks pengolahan bahasa alami (*Natural Language Processing* atau *NLP*), *case folding* biasanya mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*).

Tokenizing adalah proses memecah teks menjadi unit yang lebih kecil dan mudah dianalisis oleh komputer, disebut *token*. *Token* bisa berupa karakter individual, kata, atau bahkan frasa. Sehingga, proses ini akan melakukan pemisahan pada teks ulasan pada dokumen yang diubah menjadi suatu potongan kata yang tidak saling berpengaruh atau independen. *Text normalization* adalah proses mengubah kata menjadi bentuk standarnya. Ini bertujuan untuk membuat konsistensi dalam bagaimana program komputer memahami dan memproses teks.

Filtering (stopword removal) dalam *NLP (Natural Language Processing)* adalah proses membuang kata-kata yang dianggap tidak informatif dari teks. Kata-kata ini sering disebut *stopwords*. *Filtering* merupakan tahapan yang dimana pada tahapan ini menghapuskan kata pada ulasan yang tidak memiliki makna. Setelah itu dalam proses *filtering*, menggunakan *stopword* bahasa Indonesia yang didapatkan dari *library NLTK* pada *filtering* suatu *dataframe*. *Stemming* merupakan tahapan dalam pemrosesan teks yang bertujuan untuk menghilangkan afiks atau akhiran kata dalam teks dengan bertujuan untuk menghasilkan bentuk kata dasar dari suatu kata. Berikut merupakan salah satu contoh dalam proses *preprocessing* dalam penelitian ini yang ditunjukkan pada Tabel 2.

Tabel 2: Proses *Preprocessing* Data

Tahapan	Ulasan
Data Awal	Impulsif tbtb nonton everything everywhere all at once padahal pas nonton trailernya mikir ni film gajelas, ternyata seru bgttt dan gambaran universe nya lebih dapet daripada universe di ds mom awkwkw
<i>Case Folding</i>	impulsif tbtb nonton everything everywhere all at once padahal pas nonton trailernya mikir ni film gajelas ternyata seru bgttt dan gambaran universe nya lebih dapet daripada universe di ds mom awkwkw
<i>Tokenizing</i>	impulsif tbtb nonton everything everywhere all at once padahal pas nonton trailernya mikir ni film gajelas ternyata seru bgttt dan gambaran universe nya lebih dapet daripada universe di ds mom awkwkw
<i>Text normalization</i>	impulsif tibatiba menonton everything everywhere all at once padahal pas menonton trailernya mikir ini film enggak jelas ternyata seru banget dan gambaran universe nya lebih dapat daripada universe di ds mom awkwkw
<i>Filtering (stopword removal)</i>	impulsif tibatiba menonton everything everywhere all at once pas menonton trailernya mikir film seru banget gambaran universe nya universe ds mom awkwkw
<i>Stemming</i>	impulsif tibatiba tonton everything everywhere all at once pas tonton trailer mikir film seru banget gambar universe nya universe ds mom awkwkw

Hasil <i>Preprocessing</i>	impulsif tonton everything everywhere all at once pas tonton trailer mikir film seru banget gambar universe nya universe ds mom awkkwk
-------------------------------	---

3.3 Pelabelan Data

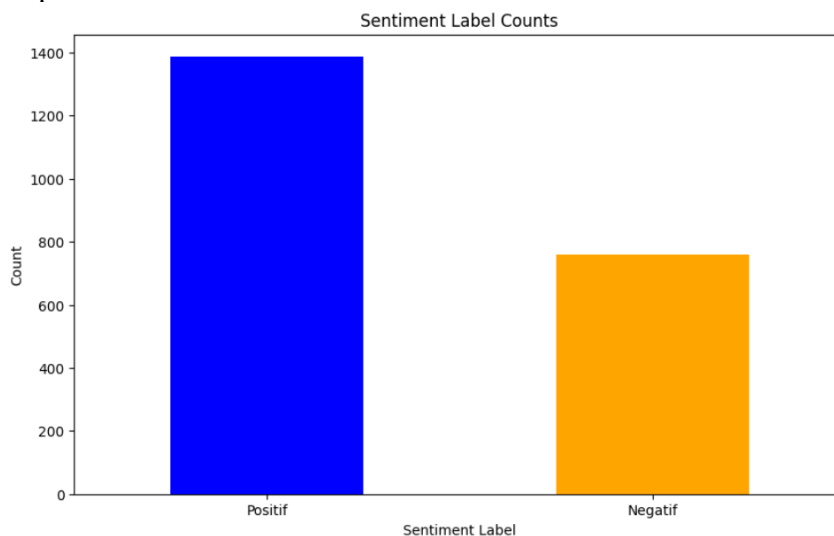
Setelah melakukan *preprocessing* data, selanjutnya yaitu dalam penelitian ini akan melakukan pelabelan data. Untuk pelabelan, digunakan kamus berbahasa Inggris dengan menginstal package NLTK dan mengimpor kelas *SentimentIntensityAnalyzer*. Penelitian ini membagi data menjadi dua label: positif dan negatif, menggunakan package “*punkt*” dan “*vader_lexicon*” untuk memecah teks menjadi kata-kata dan menganalisis sentimen dalam bahasa Inggris. Karena NLTK berbasis bahasa Inggris, *dataset* harus diterjemahkan terlebih dahulu. Untuk itu, digunakan package *googletrans* untuk mengubah teks ke bahasa Inggris.

	tweet_bersih	translated_tweet	sentiment_label
0	jatuh cinta filmfilm kayak film everything eve...	fell in love with filmfilm like film everythin...	Positif
1	vibes nya kayak habis tonton everything everyw...	the vibes are like you've watched everything e...	Positif
2	kayak keinginan cicip tipe hidup plem every...	like the desire to taste the type of life, fil...	Positif
3	jatuh cinta filmfilm film everything everywher...	fall in love film film film everything everywh...	Positif
4	sih gambar kak pau turut erudite she seems kno...	Pau's picture is also erudite, she seems to kn...	Positif
...	...	...	...
2143	festivewind broker amp everything everywhere a...	festivewind broker amp everything everywhere a...	Positif
2144	habis ajak pacar tonton everything everywhere ...	After taking my girlfriend to watch everything...	Negatif
2145	everything everywhere all at once nih hilang b...	everything everywhere, all at once, I lost the...	Positif
2146	looks biar masuk insta story gue pamer kurus g...	looks, let me enter my insta story, I show off...	Positif
2147	trcmplies top gun maverick tonton ngeringeri se...	trcmplies top gun maverick watch it's scary goo...	Negatif

2148 rows × 3 columns

Gambar 1. Hasil Pelabelan Data

Berdasarkan dari jumlah data penonton film “*Everything Everywhere All at Once*” yang terdapat dalam penelitian ini sebanyak 2138 data tweet akan dibagi untuk data yang diklasifikasikan label positif dan label negatif. Hasil proses pelabelan data diatas pembagian data untuk label positif yaitu sebanyak 1388 data dan untuk label negatif yaitu sebanyak 760 data. Berikut adalah hasil pembagian label untuk kelas positif dan kelas negatif yang ditampilkan pada Gambar 2.



Gambar 2. Grafik Hasil Pembagian Label Kelas Positif dan Label Kelas Negatif

3.4 Klasifikasi *Naïve Bayes*

Berikut merupakan langkah-langkah dalam melakukan menentukan label kelas untuk ulasan negatif dan positif dengan menggunakan perhitungan *naïve bayes*. Pada perhitungan manual ini dikarenakan data yang digunakan banyak sekali maka peneliti mengambil sampel data saja dalam perhitungan manual ini.

Tabel 3: Data Sampel (Data *training*)

Data	Tweet	Label
1	senang tonton film everything everywhere all at once bagus banget gila	Positif
2	everything everywhere all at once award bangga merinding n haru banget tonton film bagus	Positif
3	gapaham maksud film everything everywhere all at once bingung nyeritain apa sih pliss	Negatif
4	everything everywhere all at once sedih ya besok pagi mata gue bengkak inimah	Negatif
5	everything everywhere all at once is a masterpiece gila bagus banget gangotak kata orang	Positif
6	moviemenfes gua kali lihat film everything everywhere all at once sumpah film mindblowing tambah akting bagus banget recommended	Positif
7	gila everything everywhere all at once nyapu oscar nih film fresh banget sih kemas selamat deh keren	Positif
8	sesuai nama everything everywhere all at once borong nominasi oscar bagus tuh film	Positif
9	kasihan orang tonton everything everywhere all at once	Negatif
10	bikin film aneh banget anjir everything everywhere all at once film aneh	Negatif
11	kalo everything everywhere all at once menang oscar gue marah banget sih jujur bikin film benar masterpiece parah	Positif
12	tonton everything everywhere all at once menitmenit kayak bingung lamalama enjoy sampai bagus banget woiii	Positif
13	kalo film everything everywhere all at once tidur gue malas lanjut	Negatif
14	academy awards alias oscar gelar film everything everywhere all at once hasil boyong tujuh harga daftar lengkap menang oscar	Positif
15	kemarin gue tonton everything everywhere all at once tidur deh orang tuh tonton film tidur gue rumah bioskop tidur mulu tonton bete deh	Negatif
16	everything everywhere all at once film absurd film cerita hidup super absurd haru	Negatif
17	kayak salah nih tonton everything everywhere all at once	Negatif
18	puas bahagia haru lihat hasil oscar sesuai prediksi everything everywhere all at once lumayan panen very well deserved	Positif
19	jangan tonton everything everywhere all at once tau tonton bosan	Negatif
20	wow gila film everything everywhere all at once deserve oscar sih seru banget film absurd gak bisa tidur	Positif
21	kelar tonton film everything everywhere all at once bingung jalan cerita	Negatif
22	kemarin bilang everything everywhere all at once jelek	Negatif
23	orang ngantuk tonton everything everywhere all at once malah kira bagus	Negatif

Tabel 3 merupakan sampel data yang digunakan untuk melakukan perhitungan manual *naïve bayes*. Terdapat 23 data sampel yang di labelin dengan 11 data positif dan 12 data negatif. Data *training* diatas di gunakan untuk menentukan label kelas untuk tweet negatif dan positif yang belum diketahui labelnya dengan menggunakan perhitungan manual *naïve bayes*. Untuk data testing diambil beberapa contoh sampel, sebagai berikut:

Tabel 4: Data *testing*

Data	Tweet	Label
1	everything everywhere all at once bagus banget	?

Berdasarkan pada table data sampel diatas terdapat 11 kelas untuk label positif dan 12 kelas untuk label negatif dari jumlah data contoh sampel sebanyak 23 data, sehingga menghitung peluang untuk kelas positif dan kelas negatif sebagai berikut.

$$P(\text{label}|\text{positif}) = \frac{11}{23} = 0,47$$

$$P(\text{label}|\text{negatif}) = \frac{12}{23} = 0,53$$

Selanjutnya menghitung probabilitas data nomor 1 di table data testing, sebagai berikut.

Tabel 5: Probabilitas setiap kata pada tweet nomor 1

“Everything Everywhere All at Once bagus banget”	
$P(\text{"positif"} \text{"everything"}) = \frac{11}{11} = 1$	$P(\text{"negatif"} \text{"everything"}) = \frac{12}{12} = 1$
$P(\text{"positif"} \text{"everywhere"}) = \frac{11}{11} = 1$	$P(\text{"negatif"} \text{"everywhere"}) = \frac{12}{12} = 1$
$P(\text{"positif"} \text{"all"}) = \frac{11}{11} = 1$	$P(\text{"negatif"} \text{"all"}) = \frac{12}{12} = 1$
$P(\text{"positif"} \text{"at"}) = \frac{11}{11} = 1$	$P(\text{"negatif"} \text{"at"}) = \frac{12}{12} = 1$
$P(\text{"positif"} \text{"once"}) = \frac{11}{11} = 1$	$P(\text{"negatif"} \text{"once"}) = \frac{12}{12} = 1$
$P(\text{"positif"} \text{"bagus"}) = \frac{6}{11} = 0,54$	$P(\text{"negatif"} \text{"bagus"}) = \frac{1}{12} = 0,083$
$P(\text{"positif"} \text{"banget"}) = \frac{8}{11} = 0,72$	$P(\text{"negatif"} \text{"banget"}) = \frac{1}{12} = 0,083$

Berdasarkan hasil perhitungan tersebut maka dapat menghitung probabilitas dengan naïve bayes pada tweet untuk mengetahui data tersebut termasuk kelas positif atau kelas negatif. Berikut perhitungannya.

$$\begin{aligned}
 P(\text{label}|\text{positif}) &= P(\text{label}|\text{positif}) \times (P(\text{tweet} = \text{"everything"}) \times P(\text{tweet} = \\
 &\quad \text{"everywhere"}) \times P(\text{tweet} = \text{"all"}) \times P(\text{tweet} = \text{"at"}) \times P(\text{tweet} = \\
 &\quad \text{"once"}) \times P(\text{tweet} = \text{"bagus"}) \times P(\text{tweet} = \text{"banget"})) \\
 &= 0,47 \times (1 \times 1 \times 1 \times 1 \times 1 \times 0,54 \times 0,72) \\
 &= 0,182736
 \end{aligned}$$

$$\begin{aligned}
P(\text{label}|\text{negatif}) &= P(\text{label}|\text{negatif}) \times (P(\text{tweet} = \text{"everything"}) \times P(\text{tweet} = \\
&\quad \text{"everywhere"}) \times P(\text{tweet} = \text{"all"}) \times P(\text{tweet} = \text{"at"}) \times P(\text{tweet} = \\
&\quad \text{"once"}) \times P(\text{tweet} = \text{"bagus"}) \times P(\text{tweet} = \text{"banget"})) \\
&= 0,53 \times (1 \times 1 \times 1 \times 1 \times 1 \times 0,083 \times 0,083) \\
&= 0,00365117
\end{aligned}$$

Berdasarkan hasil perhitungan manual dengan *naïve bayes* diatas, dapat diketahui bahwa pada data untuk nomor 1 di table data *testing*. Hasil perhitungan tersebut menghasilkan nilai probabilitas untuk kelas positif lebih besar dengan nilai 0,182736 dan kelas negatif dengan nilai 0,00365117. Sehingga dapat disimpulkan bahwa pada hasil perhitungan manual pada data nomor 1 berdasarkan table data *testing* termasuk pada kelas positif.

3.5 Evaluasi

Sebelum mengevaluasi hasil klasifikasi, data akan dibagi menjadi data latih dan data uji dengan rasio 80%:20%, di mana 1718 data digunakan sebagai data latih dan 480 data sebagai data uji. Proses evaluasi model dilakukan untuk mengukur tingkat akurasi dari metode yang digunakan dalam penelitian ini. Hasil akurasi diperoleh melalui *confusion matrix* yang dihasilkan menggunakan Python, seperti yang ditunjukkan pada Gambar 3.

```

-----Confusion matrix:-----
      [[127  10]
       [104 189]]
-----Classification report:-----
              precision    recall  f1-score   support

   Negatif         0.55         0.93         0.69         137
    Positif         0.95         0.65         0.77         293

 accuracy          0.73
 macro avg          0.75         0.79         0.73         430
weighted avg          0.82         0.73         0.74         430

```

Gambar 3. Hasil Evaluasi Model dengan Confusion Matrix

Berdasarkan hasil yang telah ditampilkan dengan menggunakan metode *confusion matrix* pada Gambar 3. Diketahui bahwa dengan jumlah data uji sebanyak 306 data, didapatkan *true positive* 127, *true negative* 189, *false positive* 104 dan *false negative* 10. Sehingga dapat menghitung akurasi dengan persamaan pada rumus akurasi *confusion matrix*, yaitu.

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{127 + 189}{(127 + 189 + 104 + 10)} = \frac{316}{430} = 0,7348 \approx 0,73$$

Berdasarkan hasil tersebut didapatkan hasil akurasi evaluasi model dengan *confusion matrix* yaitu sebesar 87%.

3.6 Visualisasi

Tahapan berikutnya dalam penelitian ini adalah menampilkan visualisasi dalam bentuk wordcloud untuk menganalisis data ulasan dari film “*Everything everywhere all at once*” di twitter. Visualisasi ini akan memperlihatkan perbedaan antara ulasan dengan sentimen positif dan negatif, seperti yang ditunjukkan pada Gambar 3.

- (accessed Jul. 12, 2024).
- [5] F. Febriant, H. Christy, A. Wijaya, S. Informasi, U. Katolik, and M. Charitas, “Analisis Sentimen Film The Marvels Dari Aplikasi Twitter Menggunakan Metode Lexicon Based,” 2023.
 - [6] R. I. Pristiyanti, M. A. Fauzi, and L. Muflikhah, “Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor,” vol. 2, no. 3, pp. 1179–1186, 2018.
 - [7] A. S. Pangestu, H. M. Jumhur, and M. Hum, “PENGARUH RATING IMDB TERHADAP KEPUTUSAN MENONTON FILM GAME OF THRONES DI KOTA BANDUNG TAHUN 2018 IMPACT OF IMDB RATING ON THE DECISION TO WATCH GAME OF THRONES FILM IN BANDUNG CITY IN 2018,” vol. 5, no. 3, pp. 3253–3258, 2018.
 - [8] R. Amelia, N. S. Prastiwi, and M. E. Purbaya, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter,” vol. 9, no. 2, pp. 338–343, 2022, doi: 10.30865/jurikom.v9i2.3895.
 - [9] O. U. Effendy, *Dinamika Komunikasi*. 2008.
 - [10] H. Oktavianus, “Prakter Eksorsis di dalam Film Conjuring,” *E-Komunikasi Univ. Kristen Petra Surabaya*, vol. 3, no. 2, pp. 1–12, 2015, [Online]. Available: <https://media.neliti.com/media/publications/79600-ID-none.pdf>
 - [11] S. Azeharie and O. Kusuma, “Analisis Penggunaan Twitter Sebagai Media Komunikasi Selebritis Di Jakarta,” pp. 83–98, 2014.
 - [12] N. M. and N. K. W. Sriyanong, “A Text Preprocessing Framework for Text Mining on Big Data Infrastructure,” *2018 2nd Int. Conf. Imaging, Signal Process. Commun.*, pp. 169–173, 2018.
 - [13] Z. R. and R. A. P. U. Hasanah, T. Astuti, R. Wahyudi, “An Experimental Study of Text Preprocessing Techniques for Automatic Short Answer Grading in Indonesian,” *2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng.*, pp. 230–234, 2018.
 - [14] Arifin Kurniawan, Indriati Indriati, and Sigit Adinugroho, “Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 9, pp. 8335–8342, 2019.
 - [15] Khaeva, D. N., & Basir, C. (2024). Implementasi Metode Convolutional Neural Network (CNN) Untuk Mendeteksi Jenis Tanaman Herbal. *Jurnal SAINTEK Patompo*, 2, 65–73.