

ANALISIS SENTIMEN ULASAN INSTAGRAM DI GOOGLE PLAY STORE MENGUNAKAN ALGORITMA NAÏVE BAYES

Dhodi Surya Sayogo, Bambang Irawan, Agus Bahtiar

Program Studi Teknik Informatika S1

STMIK IKMI CIREBON. Jl. Perjuangan No.10B, Karyamulya, Kota Cirebon

dhodisuryasayogo@gmail.com

ABSTRAK

Instagram sebuah platform media sosial yang populer, telah mengubah lanskap komunikasi dan interaksi sosial dalam era digital. Dalam abstrak ini, menyelidiki perkembangan, dampak, dan peran *Instagram* dalam kehidupan sehari-hari penggunaannya. *Instagram*, yang diluncurkan pada tahun 2010, fokus pada berbagi foto dan video dengan berbagai filter kreatif, memberikan pengguna kemampuan untuk mengekspresikan diri secara visual. Teknik *crawling* untuk mengumpulkan ulasan aplikasi dari *Google Play Store* dengan metode klasifikasi *Naïve Bayes* untuk menganalisis sentimen ulasan. Metode klasifikasi *Naïve Bayes* digunakan untuk mengklasifikasikan ulasan sebagai positif dan negatif berdasarkan bahasa dan kata-kata yang digunakan dalam ulasan tersebut. Analisis sentimen adalah bidang yang berkembang pesat dalam pengolahan bahasa alami dan data mining. Ini melibatkan pengidentifikasian dan klasifikasi opini, perasaan, atau sentimen yang terkandung dalam teks, sering kali dengan tujuan memahami pandangan atau reaksi publik terhadap berbagai topik, produk, atau layanan. Dalam abstrak ini, Tujuan penelitian ini adalah untuk menganalisis ulasan user terhadap *Instagram* yang berada di *Google Play Store*. Metode yang digunakan dalam penelitian ini menggunakan teknik *Naïve Bayes*. Dataset berisi tentang ulasan mengenai aplikasi *instagram* di *Google Play Store* sebanyak 1100 dataset. Hasil penelitian ini menunjukkan *accuracy* 89%, *precision* 90% dan *recall* 98%.

Kata kunci: Analisis Sentimen, *Crawling*, *Google Play Store*, *Instagram*, *Naïve Bayes*

1. PENDAHULUAN

Kehadiran Internet bisa dibilang terlambat di Indonesia, namun perkembangannya pun sangat cepat. Pada awalnya Internet hanya tersedia di media komputer, namun saat ini sudah ada di media elektronik yang kita bisa bawa kemana mana seperti tablet, laptop dan handphone. Saat ini media media konvensional seperti televisi, radio dan koran telah dikesampingkan oleh masyarakat di dunia, termasuk di Indonesia [1].

Penelitian ini akan mengimplementasikan metode data *crawling* untuk mengumpulkan datanya yang bertujuan untuk menganalisa tanggapan publik mengenai Instagram dan mengklasifikasikannya menjadi dua bagian yaitu positif dan negatif. Hasil dari Analisa sentimen ini dapat menjadi pertimbangan developer untuk membantu dalam pengembangan dan peningkatan aplikasi. Penelitian ini memiliki dampak positif dalam meningkatkan kualitas layanan, meningkatkan pengalaman pengguna, dan memberikan wawasan yang berharga bagi berbagai pemangku kepentingan di ekosistem Instagram dan *Google Play Store*.

Berdasarkan penelitian terdahulu yang pernah dilakukan oleh [2]. mengenai Analisis Sentimen Pelanggan *Tokopedia* yang menggunakan algoritma *Naïve Bayes Classifier*. Performa yang dihasilkan pada pengujian yang dilakukan oleh *rapid evaluator* pada 143 data pengujian menghasilkan nilai akurasi 95,10%. Analisis ini menggunakan metode *Naïve Bayes Classifier* terhadap komentar pelanggan tentang aplikasi *Tokopedia* di *Playstore*. Pada penelitian selanjutnya [3] ulasan mengenai aplikasi *flip* diambil

sebanyak 200 dataset yang dibagi menjadi 100 data positif dan 100 data negatif diolah menggunakan *rapid miner* dengan menerapkan metode *naïve bayes* yang dioptimasi dengan PSO. Hasil pengolahan menggunakan *naïve bayes* 82,00%, kemudian di optimasi dengan PSO dan didapatkan hasil 88,24%.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Bobby Kurniadi Widodo, Nur Hafifah Matondang, dan Desta Sandya Prasvita penelitian yang berjudul Penerapan Algoritma *Naïve Bayes* Untuk Analisis Sentimen Penggunaan Aplikasi *Jobstreet*. Tujuan dari penelitian ini adalah melakukan klasifikasi sentimen terhadap ulasan pada aplikasi *Jobstreet* dengan metode *Naïve Bayes*. Dalam penelitian ini opini akan dibagi kedalam dua golongan sebagai positif dan negatif, kemudian diklasifikasikan dengan menggunakan algoritma *Naïve Bayes*. Hasil pengujian yang didapat menggunakan data uji memiliki nilai akurasi sebesar 0,96; nilai precision sebesar 0,98; nilai recall sebesar 0,94 [4].

Penelitian selanjutnya oleh Ernianti Hasibuan dan Elmo Allistair Heriyanto penelitian yang berjudul Analisis Sentimen Pada Ulasan Aplikasi *Amazon Shopping* Di *Google Play Store* Menggunakan *Naïve Bayes Classifier*. Pada penelitian ini akan dilakukan analisis sentimen dengan pendekatan berbasis *machine learning* menggunakan algoritma *Naïve Bayes* menggunakan data ulasan pengguna pada aplikasi *Amazon Shopping* yang terdapat pada *Google Play Store*. Hasil klasifikasi menggunakan keempat algoritma *Naïve Bayes* menghasilkan akurasi rata-rata

sebesar 82.15%, precision sebesar 72.25%, recall sebesar 83.49%, dan f1-score sebesar 77.41%. Multinomial NB menghasilkan akurasi terbaik diantara keempat algoritma *Naive Bayes* yang digunakan, yaitu sebesar 86.74%. Nilai *precision*, *recall*, dan *f1-score* berturut-turut adalah 78.82%, 85.90%, dan 82.21% [5].

Penelitian selanjutnya Hindun Habibatul Mubaroroh, Hasbi Yasin, dan Agus Rusgiyono dengan judul penelitian Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs *Google Play* Menggunakan Algoritma *Naive Bayes Classifier* Dengan Normalisasi Kata Levenshtein Distance. Metode yang digunakan *10-fold cross validation* karena mudah digunakan dengan struktur yang sederhana dan memberikan nilai akurasi yang tinggi. Tahap normalisasi kata juga disempurnakan menggunakan metode Levenshtein Distance yang terbukti menambah nilai akurasi. Hasil performa menggunakan Multinomial *Naive Bayes* dengan menambahkan metode Levenshtein Distance untuk memperbaiki kata memberikan nilai akurasi rata-rata sebesar 88,20% dengan lipatan ke-8 sebagai lipatan dengan nilai akurasi terbaik sebesar 94% [6].

Penelitian selanjutnya Bintang Zulfikar Ramadhan, Ibnu Riza, dan Iqbal Maulana dengan judul Analisis Sentimen Ulasan pada Aplikasi *E-Commerce* Dengan Menggunakan Algoritma *Naive Bayes*. Penelitian ini menggunakan metode *Knowledge Discovery in Database (KDD)* yang terdiri dari 5 tahap yaitu seleksi data, *preprocessing*, transformasi, data mining, dan evaluasi. Data yang digunakan adalah *review* terhadap 500 aplikasi *E-Commerce* per masing-masing aplikasi. Pada tahap data mining dilakukan dengan 3 skenario pembagian data yaitu 80:20, 70:30 dan 60:40. Hasil terbaik diperoleh pada skenario 1 (80:20) pada aplikasi *Shopee* dengan menggunakan algoritma *Naive Bayes* yang menghasilkan akurasi sebesar 92%, presisi sebesar 92,13%, *recall* sebesar 98,8% dan *f1-score* sebesar 95,35% [7].

Penelitian selanjutnya Rita Apriani dan Dudih Gustian dengan judul Analisis Sentimen dengan *Naive Bayes* Terhadap Komentar Aplikasi *Tokopedia*. Penelitian ini menggunakan metode *Naive Bayes* untuk menghasilkan sentimen positif dan negatif terhadap komentar pengguna aplikasi *Tokopedia* di *Playstore*. Pengujian berdasarkan nilai *class negative*, *class positive*, *recall*, dan *accuracy* pada analisis sentimen. dengan nilai *accuracy performance* yang baik sebesar 97,13%, dengan nilai *precision* 1 Sementara pada *Class Recall* dihasilkan nilai 95,49% (*positive class: negative*). Dan nilai AUC 0,980 [8].

2.2. Text Mining

Text Mining merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi, dimana *text mining* merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [9].

Jenis data ini dapat berasal dari berbagai sumber, seperti platform media sosial, artikel berita, ulasan pelanggan, atau makalah penelitian [10].

2.2. Analisis Sentiment

Analisis sentimen, yang juga disebut dengan *opinion mining* merupakan bidang studi yang menganalisis pendapat, penilaian, evaluasi, sentimen, sikap dan emosi terhadap entitas seperti produk, jasa, organisasi, individu, peristiwa, topik dan atribut lainnya. Sejak awal tahun 2000, analisis sentimen telah berkembang menjadi salah satu penelitian aktif dalam *natural language processing*. Dalam beberapa tahun terakhir, analisis sentimen pada kegiatan industri juga terus berkembang [11].

2.3. Algoritma Naive Bayes

Naive Bayes adalah salah satu algoritma pembelajaran mesin dan data mining yang paling efektif dan efisien. Meskipun asumsi bahwa atribut dalam data adalah independen, kinerja klasifikasi *Naive Bayes* tetap cukup tinggi. Meskipun asumsi independensi atribut jarang terjadi dalam data yang sebenarnya, jika dilanggar, algoritma tetap dapat menghasilkan hasil klasifikasi yang baik [12]. Model dari *Naive Bayes* yang sering digunakan dalam klasifikasi teks adalah *Multinomial Naive Bayes* yang memperhitungkan frekuensi dari setiap kata yang muncul pada setiap ulasan [6].

2.4. Teknik Crawling

Teknik *Crawling* adalah suatu teknik yang digunakan untuk mengumpulkan informasi yang ada dalam web [13]. Hasil dari *crawling* data disimpan dalam file bertipe *.csv* dan kemudian dilakukan labelling untuk menentukan pendapat atau pandangan dari *review* yang diambil. Pada proses pelabelan dibedakan menjadi 2 kelas, yaitu kelas positif dan kelas negatif. [14].

2.5. Evaluasi Akurasi

Untuk mengukur akurasi dari setiap metode sebelumnya maka akan diukur menggunakan nilai *accuracy*, *precision*, dan *recall*. Terdapat empat istilah sebagai representasi hasil proses klasifikasi. Berikut tabel dari empat istilah tersebut [15]:

Tabel 1. *Confusion matrix*

	Predicted Class	
True Class	TP	FN
	FP	TN

Penjabaran tabel diatas sebagai berikut :

- TP (*True Positif*) : Kumpulan data positif yang terdeteksi positif.
- TN (*True Negative*) : Kumpulan data positif yang terindikasi negatif.
- FP (*False Positive*) : Kumpulan data negatif terindikasi positif.

- FN (False Negative) : Kumpulan data negatif terindikasi negatif.

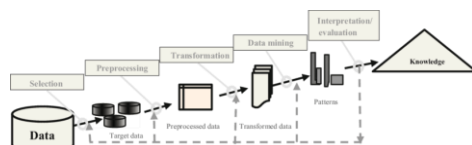
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

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

3. METODE PENELITIAN

Metode penelitian merupakan suatu pendekatan yang digunakan untuk menguji sebuah asumsi. Dalam strategi eksperimental terhadap sentimen, teks-teks diklasifikasikan ke dalam kategori positif dan negatif sebelum dianalisis guna menentukan tingkat sentimen yang terkandung di dalamnya. Proses ini memanfaatkan metode *Naïve Bayes* untuk mengukur *accuracy*, *precision*, dan *recall* dari algoritma tersebut.



Gambar 1. Knowledge discovery in data (KDD)

3.1. Selection

Tahap awal pengambilan data yang berdasarkan ulasan relevan dari user pengguna *instagram*. Data diambil menggunakan *API google-play-scraper* dengan teknik crawling dan mendapatkan 1100 ulasan yang ditulis oleh user pengguna *instagram*. Melakukan EDA (*Exploraty Data Analysis*) untuk melihat data berdasarkan score dari ulasan pengguna *instagram*. Selanjutnya data tersebut diambil 2 variable *content* dan *score* dan di tambahkan variable label yang berfungsi sebagai kolom hasil dari score tersebut apakah negatif atau positif.

3.2. Preprocessing

Pembersihan data untuk proses mengidentifikasi bagian data yang tidak lengkap, salah, benar, tidak akurat atau hilang dan mengoreksi, mengganti atau menghapusnya. Cleansing penelitian ini akan menghapus @, #, karakter *URL* serta attribut yang hanya menggunakan teks.

3.3. Transformation

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Di tahap ini peneliti menggunakan berbagai macam library untuk mengubah kata dari mulai mengubah semua karakter dalam sebuah teks menjadi huruf kecil atau huruf besar menggunakan library *re* (*RegEx*) atau *Regular Expressions*, menghilangkan kata memiliki makna yang signifikan, seperti "a", "an",

"the", "in", dan sebagainya menggunakan library *stopwords*, dan library lainnya.

3.4. Data Mining

Mengubah teks yang tidak terstruktur menjadi format terstruktur untuk mengidentifikasi pola yang bermakna dan wawasan baru. Pada tahap Riset ini menerapkan *Naive Bayes*.

3.5. Evaluation

Proses evaluasi dari rule yang sudah didapat sesuai dengan pola dan tahapan yang sudah ditentukan. Tahap ini juga menggunakan aplikasi *anaconda navigator* untuk memunculkan *jupyter notebook* menentukan hasil *accuracy* yang diperoleh menggunakan *Algoritma Naive Bayes*.

4. HASIL DAN PEMBAHASAN

4.1. Library Python

Sebelum menganalisis sentimen ulasan *instagram* ada beberapa library yang harus di jalankan agar menganalisis dapat berjalan dengan lancar. Berikut beberapa library yang harus diinstall :

```
In [ ]: # library untuk mengambil data ulasan di website google play store
1 | pip install google-play-scraper
2 | # berfungsi sebagai Scraper Data ulasan user dari google playstore
3 | from google_play_scraper import app
4 | # library untuk akses bahasa pemrograman python yang digunakan untuk pengolahan data.
5 | import pandas as pd
6 | # berfungsi untuk membuat array dari struktur data
7 | import numpy as np
8 | # regex digunakan untuk mencocokkan, mencari, mengganti, dan memanipulasi teks berdasarkan pola tertentu.
9 | import re
10 | # nltk berfungsi untuk pengolahan data bahasa manusia.
11 | # library nltk adalah bagian dari paket nltk yang digunakan untuk bahasa alami.
12 | from nltk.corpus import stopwords
13 | import nltk.corpus
14 | # nltk berfungsi untuk pengolahan data bahasa manusia.
15 | # bisa digunakan berfungsi untuk menghasilkan token-cmu, direpresentasikan sebagai tuple. string bulat
16 | import nltk
17 | from nltk.tokenize import sent_tokenize, word_tokenize
18 | # library abstract adalah sebuah modul sederhana yang diletakkan oleh library python yang memungkinkan untuk melakukan
19 | # pengisian kata-kata yang ter-infleksi dalam bahasa Indonesia ke bentuk baku-nya
20 | pip install Sastrawi
21 | # modul untuk cleaner yang dapat digunakan untuk melakukan cleaning pada data teks berbahasa Indonesia.
22 | from sastrawi.stemmer.StemmerFactory import StemmerFactory
```

Gambar 2. Library python yang dibutuhkan

Berikut penjelasan mengenai Gambar 2 yang yang di butuhkan diantaranya :

1. Library *google-play-scraper* berfungsi sebagai mengambil data ulasan *instagram* di *google play store*.
2. Library *pandas* digunakan untuk pengolahan data.
3. Library *numpy* berfungsi untuk membuat array dari struktur data.
4. Library *re(RegEx)* digunakan untuk mencocokkan, mencari, mengganti, dan memanipulasi teks berdasarkan pola tertentu.
5. Library *nltk* berfungsi untuk pengolahan data bahasa manusia.
6. Library *sastrawi* diubakan pengurang kata-kata yang ter-infleksi dalam bahasa Indonesia ke bentuk baku-nya.
7. Library *sklearn* membantu melakukan procesing data ataupun melakukan training data untuk kebutuhan machine learning atau data science.

4.2. Pengumpulan Data

Mengumpulkan data dengan cara *crawling* yang dimana mengambil data melalui API aplikasi *instagram* yang di dapatkan dari website *google play store*. Setelah API didapatkan langkah selanjutnya memasukan code phyton seperti di **Gambar 3** jika di jelaskan sebagai berikut, aplikasi yang berada di indonesia(id), berbahasa Indonesia(id), serta kita sortir ulasan MOST_RELEVANT bertujuan untuk mendapatkan hasil ulasan yang saling relevan(berkaitan), jumlah data yang ingin di ambil adalah 1100 data, filter data NONE agar dapat mengambil semua score/ bintang 1 sampai 5.

```

1 # scrape jumlah ulasan yang diinginkan
2 from google_play_scraper import Sort, reviews
3
4 result, continuation_token = reviews(
5     'com.instagram.android',
6     lang='id', # scrape data ulasan aplikasi yang berada di google play store
7     country='id', # setting bahasa menjadi bahasa Indonesia
8     sort=Sort.MOST_RELEVANT, # gunakan most_relevant untuk mendapatkan ulasan yang paling relevan
9     count=1100, # filter: jumlah ulasan yang mau di ambil ada 1100 data
10    filter_score_with=None # kemudian di filter_score gunakan None untuk mengambil semua score bintang 1 sampai 5
11 )
    
```

Gambar 3. Proses pengambilan data

4.3. Data Labeling

Setelah data di dapatkan melanjutkan tahapan ke labeling memberikan positif dan negatif menggunakan atribut score sebagai acuan untuk memeberikan label. pada Gambar 4 pelabelan menggunakan percabangan *if else* yang menjadi titik acuan adalah atribut score jika kurang dari 3 itu mendapatkan hasil *negatif*, jika atribut score sama dengan 4 maka akan dihasilkan positif, terakhir jika atribut sama dengan 5 maka akan dihasilkan positif. Hasil dari pelabelan akan masuk kedalam atribut label yang di buat.

```

1 # untuk memberikan Label positif atau negatif dari score
2 def pelabelan(score):
3     if score < 3:
4         return 'Negatif'
5     elif score == 4 :
6         return 'Positif'
7     elif score == 5 :
8         return 'Positif'
9 my_df['Label'] = my_df ['score'].apply(pelabelan)
10 my_df.head(100)
    
```

Gambar 4. Proses pelabelan

4.4. Atribut Data

Data yang digunakan mengenai ulasan pengguna Instagram yang ada di *google play store* dengan teknik *crawling*, menggunakan API *com.instagram.android*, dengan jumlah data 1100 ulasan pengguna *insstagram*, yang berjumlah 3 atribut sebagai berikut :

Tabel 2. Atribut data yang digunakan

No	Atribut	Tipe
1.	Content	Object
2.	Score	Int64
3.	Label	Object

Pada Tabel 1 memiliki penjelasan tipe data sebagai berikut :

1. Tipe data *object* merupakan jenis data yang tidak hanya menyimpan informasi, tetapi juga mencakup petunjuk tentang bagaimana data

tersebut dapat diolah. Secara kontras, tipe data lain seperti integer, float, dan string hanya menyimpan jenis data masing-masing tanpa menyertakan informasi tambahan atau cara pengolahan data.

2. Tipe data *integer(int)* merupakan data type berbentuk bilangan bulat atau numerik yang umumnya digunakan untuk menyimpan angka tanpa komponen pecahan dengan rentang angka -707, 0, hingga 707. Type data score adalah *int64* adalah jumlah bit yang dipakai untuk menyimpan nilai masing-masing: integer 64 bit menampung lebih banyak daripada integer 32 bit. Integer 64 memiliki rentang nilai valid yang lebih tinggi.

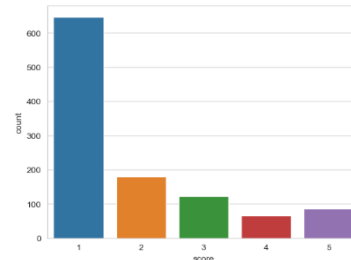
4.5. Visualisasi Chart

Selanjutnya dilakukan dua kali menampilkan *Exploraty Data Analysis* (EDA) sebelum cleaning dan sesudah cleaning, data masih lengkap rating 1,2,3,4,5 berguna untuk mengidentifikasi pola, menemukan anomali, menguji hipotesis dan memeriksa asumsi.

```

Input data has 1100 rows and 3 columns
score 1 = 647 rows
score 2 = 179 rows
score 3 = 122 rows
score 4 = 66 rows
score 5 = 86 rows
    
```

Gambar 5. Hasil atribut score



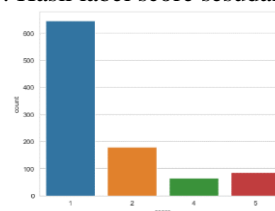
Gambar 6. Visualisasi scatterplot atribut score

Output pada Gambar 5 dan Gambar 6 adalah jumlah ulasan tiap score dan data yang belum di cleaning masih menjadi data mentah walaupun datanya sudah di label kan positif dan negatif.

```

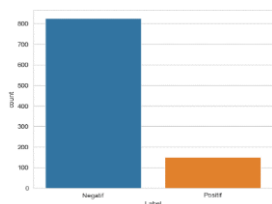
Input data has 978 rows and 3 columns
score 1 = 647 rows
score 2 = 179 rows
score 3 = 0 rows
score 4 = 66 rows
score 5 = 86 rows
    
```

Gambar 7. Hasil label score sesudah dicleaning



Gambar 8. Visualisasi scatterplot label score

Terlihat ada perubahan pada Gambar 7 dan Gambar 8 yaitu menghilangkan score 3 yang tidak terpakai bawasannya score 3 mempunyai alasan yang ambigu dalam hal ini hanya memakai score 1, 2, 4, dan 5.



Gambar 9. Hasil scatterplot atribut label

```
Negatif      826
Positif      152
Name: Label, dtype: int64
```

Gambar 10. Hasil dari atribut label

Setelah memasukan tahap cleaning dan menghapus nilai NaN pada Gambar 9 dan Gambar 10 menghasilkan yang berlabel positif 152 ulasan dan negatif 826 ulasan jika ditambah kan keseluruhan atribut label berjumlah 978 ulasan yang berlabel positif dan negatif dari 1100 dataset yang sebelumnya diambil menggunakan library *google-play-scraper*.

4.6. Hasil Pengujian

Hasil pengujian menggunakan algoritma *Naive Bayes*:

```
MultinomialNB Accuracy: 0.8979591836734694
MultinomialNB Precision: 0.9065934065934066
MultinomialNB Recall: 0.9821428571428571
MultinomialNB f1_score: 0.9428571428571428
-----
              precision    recall  f1-score   support
Negatif         0.91         0.98         0.94         168
Positif         0.79         0.39         0.52          28

accuracy         0.90         0.90         0.90         196
macro avg        0.85         0.69         0.73         196
weighted avg     0.89         0.90         0.88         196
```

Gambar 11. Hasil report

Berdasarkan hasil pengujian menggunakan algoritma *Naive Bayes* pada sentimen analisis ulasan instagram di google play store diperoleh hasil *accuracy* 89%, *precision* 90%, dan *recall* 98%.

4.7. Confusion Matrix

Setelah melakukan hasil uji dari algoritma *Naive Bayes* melanjutkan menampilkan hasil *confusion matrix* yang berguna untuk memastikan apakah hasil *accuracy*, *precision*, dan *recall* benar dilakukan menggunakan rumus, pada Gambar

```
Hasil Laporan confusion matrix Algoritme naive bayes:
[[165  3]
 [ 17 11]]
```

Gambar 12. Confusion matrix Naive Bayes

Mendapatkan hasil tabel di bawah ini :

Tabel 3. Tabel Confusion Matrix Naive Bayes

True Class	Predicted Class	
		165
	17	11

Hasil uji algoritma *Naive Bayes* selanjutnya dihitung berdasarkan rumus berikut :

$$Accuracy = \frac{11 + 165}{11 + 165 + 17 + 3} = 0,89 \quad (4)$$

$$Precision = \frac{165}{165 + 17} = 0,90 \quad (5)$$

$$Recall = \frac{165}{165 + 3} = 0,98 \quad (6)$$

5. KESIMPULAN DAN SARAN

Pada percobaan ini dapat disimpulkan analisis Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naive Bayes. Dari pengujian yang telah dilakukan menghasilkan: Hasil penelitian ini menggunakan model analisis sentimen Naive Bayes jenis Multinomial Naive Bayes yang mana model tersebut cocok untuk dipergunakan klasifikasi yang diolah berdasarkan tahapan Knowledge Discovery in Database (KDD). Peneliti hanya menggunakan score 1,2,4, dan 5 setelah membaca data di variable Label score 3 menghasilkan review yang tidak pasti yang mana hasilnya bisa positif dan negatif. Hasil dari 1100 ulasan setelah melakukan preprocessing data menghasilkan 978 ulasan yang berlabel dan dipecah berdasarkan Variable Label memiliki nilai 152 ulasan positif dan 826 ulasan negatif. Hasil penelitian ini tentang ulasan review instagram cenderung memberikan tanggapan atau opini negatif terhadap aplikasi instagram. Hasil akurasi dalam penelitian ini menggunakan algoritma Naive Bayes bernilai 89%, precision bernilai 90%, dan recall bernilai 98%.

Adapun saran yang dapat di sampaikan yaitu sebagai berikut : Diharapkan adanya tambahan dataset sehingga dapat meningkatkan jumlah akurasi dalam ulasan review instagram di google play store. Menggunakan metode lain sehingga dapat melihat setiap akurasi yang didapatkan dari proses penelitian. Penelitian selanjutnya dapat mencari sumber data lain.

DAFTAR PUSTAKA

- [1] D. N. Sari and A. Basit, "Media Sosial Instagram Sebagai Media Informasi Edukasi," *Persepsi Commun. J.*, vol. 3, no. 1, pp. 23–36, 2020, doi: 10.30596/persepsi.v3i1.4428.
- [2] S. M. Salsabila, A. Alim Murtopo, and N. Fadhillah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naive Bayes Classifier," *J. Minfo Polgan*, vol. 11, no. 2, pp. 30–35, 2022, doi: 10.33395/jmp.v11i2.11640.
- [3] O. Irnawati and K. Solecha, "Analisis Sentimen Ulasan Aplikasi Flip Menggunakan Naive Bayes

- dengan Seleksi Fitur PSO,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 189–199, 2022.
- [4] B. K. Widodo, N. H. Matondang, and D. S. Prasvita, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Penggunaan Aplikasi Jobstreet,” *Techno.Com*, vol. 21, no. 3, pp. 523–533, 2022, doi: 10.33633/tc.v21i3.6361.
- [5] Ernianti Hasibuan and Elmo Allistair Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
- [6] H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, “Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naive Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance,” *J. Gaussian*, vol. 11, no. 2, pp. 248–257, 2022, doi: 10.14710/j.gauss.v11i2.35472.
- [7] B. Z. Ramadhan, R. I. Adam, and I. Maulana, “Analisis Sentimen Ulasan pada Aplikasi E-Commerce dengan Menggunakan Algoritma Naive Bayes,” *J. Appl. Informatics Comput.*, vol. 6, no. 2, pp. 220–225, 2022, doi: 10.30871/jaic.v6i2.4725.
- [8] R. Apriani *et al.*, “Analisis Sentimen dengan Naive Bayes Terhadap Komentar Aplikasi Tokopedia,” *J. Rekayasa Teknol. Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, [Online]. Available: <https://rekayasa.nusaputra.ac.id/article/view/86>
- [9] C. Joergensen E Munthe, N. Astuti Hasibuan, and H. Hutabarat, “Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace,” *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 2, no. 3, pp. 110–115, 2022, doi: 10.30865/resolusi.v2i3.309.
- [10] M. Tegar Lazuardi, T. Suprpti, and Y. Arie Wijaya, “Perancangan Model Sentimen Tweet Terhadap Pilkada Dki Jakarta Tahun 2017 Menggunakan Algoritma Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 308–312, 2023, doi: 10.36040/jati.v7i1.6328.
- [11] F. Alvianda, “Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 1, pp. 241–246, 2019.
- [12] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, “Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [13] B. R. Atmadja, “Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata Di Kabupaten Sukabumi Dengan Naive Bayes Classifier,” *Elkom J. Elektron. dan Komput.*, vol. 15, no. 2, pp. 371–382, 2022, doi: 10.51903/elkom.v15i2.872.
- [14] T. Astuti and Y. Astuti, “Analisis Sentimen Review Produk Skincare Dengan Naive Bayes Classifier Berbasis Particle Swarm Optimization (PSO),” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 1806, 2022, doi: 10.30865/mib.v6i4.4119.
- [15] M. M. Ismail and K. M. Lhaksamana, “Sentimen Analisis Pada Media Online Mengenai Pemilihan Presiden 2019 Dengan Menggunakan Metode Naive Bayes,” *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9997–10004, 2019.