
ANALISIS SENTIMEN PADA ULASAN APLIKASI AIRBNB DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE BAYES CLASSIFICATION

Elis Mulyani¹, Eko Martantoh², Purwanto³

^{1,2,3}Universitas Panca Sakti Bekasi

Email: elismulyani0507@gmail.com

Abstrak: Aplikasi Airbnb merupakan wadah bagi komunitas pemasaran yang terpercaya atau aplikasi yang bisa membuat siapa saja untuk memasarkan tempat yang disewakannya, menelusuri, dan memesan tempat akomodasi yang unik di seluruh penjuru dunia via aplikasi online melalui ponsel pribadi (Airbnb, 2015). Beberapa akomodasi yang ada pada aplikasi Airbnb diantaranya ada apartemen, kastil, vila, kamar, dan fasilitas akomodasi lainnya serta ruangan-ruangan yang dapat disewakan lainnya. Mirip dengan online travel agent lainnya, yang membedakan adalah Airbnb menyewakan tempat tinggal berupa rumah, apartemen dan villa. Data yang digunakan pada penelitian ini merupakan data sentimen dari pengguna aplikasi android yang memberikan komentarnya pada fasilitas user review pada website googleplay yang kemudian dipisahkan menjadi sentimen positif dan sentimen negatif. Dataset didapatkan dengan melakukan *scraping* di *Play Store* pada aplikasi pengguna AirBnB didapatkan 500 data ulasan berbagai pengguna yang nantinya dijadikan bahan pengujian. Hasil pengujian menggunakan metode multinomial *Naive Bayes* yang dilakukan dengan data (20% dari dataset secara *random*) pada aplikasi AirBnB digunakan berjumlah 89 data didapatkan hasil *confusion matriks* TP 33, TN 46, FP 6 dan FN 4. Dengan nilai Precision sebesar 0.84615 (85%), nilai Recall 0.89189 (89%), nilai F1-score sebesar 0.86842 (87%) dan Nilai tingkat keakuratan algoritma sebesar 0.88764 atau sekitar 89%.

Kata Kunci: Airbnb, Analisis Sentimen, Naïve Bayes, Play Store.

Abstract: The Airbnb application is a platform for a trusted marketing community or application that can enable anyone to market the place they rent, browse and book unique accommodations throughout the world via an online application via personal cell phone (Airbnb, 2015). Some of the accommodations available on the Airbnb application include apartments, castles, villas, rooms and other accommodation facilities as well as other rooms that can be rented out. Basically, Airbnb is an application with the theme of renting accommodation in the form of rooms. Similar to other online travel agents, the difference is that Airbnb rents out residences in the form of houses, apartments and villas. The data used in this research is sentiment data from Android application users who provide comments on the user review facility on the GooglePlay website which is then separated into positive sentiment and negative sentiment. The dataset was obtained by scraping the Play Store on the AirBnB user application, obtaining 500 review data from various users which would later be used as testing material. The test results using the multinomial Naive Bayes method which was carried out with data (20% of the dataset randomly) on the AirBnB application used amounted to 89

data, resulting in confusion matrices TP 33, TN 46, FP 6 and FN 4. With a precision value of 0.84615 (85%), the Recall value is 0.89189 (89%), the F1-score value is 0.86842 (87%) and the algorithm accuracy level value is 0.88764 or around 89%.

Keywords: *AirBnB, Sentiment Analysis, Naïve Bayes, Play Store.*

PENDAHULUAN

Kemajuan teknologi informasi berkembang sangat pesat hingga saat ini dimana semua aspek kebutuhan manusia menginginkan sesuatu yang serba cepat dan praktis, di mulai dari pendidikan, perdagangan, pemerintahan hingga komunikasi sosial. Perkembangan pesat tersebut tidak dapat dicegah oleh siapapun karena kebutuhan dan masalah manusia yang sangat beragam akan memicu setiap orang untuk membuat teknologi yang dapat menyelesaikan suatu masalah. Salah satu contoh teknologinya yaitu sistem pemesanan (booking) secara online. Sistem booking online merupakan cara yang mudah untuk mencari, memesan barang atau jasa melalui internet. “Informasi mengenai booking online semuanya disimpan secara digital dalam sistem komputer milik perusahaan” (Sentosa, 2015). Adapun penggunaan booking online dapat mempermudah kegiatan operasional pada suatu perusahaan, tidak terkecuali pada E-Booking jasa pemesanan akomodasi berbasis aplikasi yakni “*AirBnB*”.

Aplikasi Airbnb merupakan wadah bagi komunitas pemasaran yang terpercaya atau aplikasi yang bisa membuat siapa saja untuk memasarkan tempat yang disewakannya, menelusuri, dan memesan tempat akomodasi yang unik di seluruh penjuru dunia via aplikasi online melalui ponsel pribadi (Airbnb, 2015). Beberapa akomodasi yang ada pada aplikasi Airbnb diantaranya ada apartemen, kastil, vila, kamar, dan fasilitas akomodasi lainnya serta ruangan-ruangan yang dapat disewakan lainnya. Keberadaan Airbnb sebagai pendatang baru dalam dunia bisnis membuat ramai dunia pariwisata. Pada dasarnya Airbnb adalah sebuah aplikasi yang bertemakan penyewaan akomodasi berbentuk kamar. Mirip dengan online travel agent lainnya, yang membedakan adalah Airbnb menyewakan tempat tinggal berupa rumah, apartemen dan villa. Bentuk Airbnb sendiri merupakan bentuk Peer-to-Peer (P2P) yaitu komunikasi dua arah antara penjual dan pemilik layanan, memberikan lahan untuk penjual menjual produk mereka, pemilik atau pemberi lahan yang dimaksud disini adalah Airbnb. Airbnb menawarkan berbagai pilihan tempat untuk akomodasi dari para host (pemilik rumah/tempat tinggal) yang sudah mendaftarkan tempat tinggalnya di Airbnb dengan menarik pelanggan dari situs online ke toko fisik. Konsep Airbnb sendiri dikenal sebagai konsep

sharing economy atau juga dikenal sebagai collaborative consumption menurut (Oskam dan Boswijk, 2016)

Naive bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (independen). Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasian probabilitas. Keuntungan algoritma naive bayes adalah tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi naive bayes dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang jumlahnya lebih besar.

Dengan demikian, Penulis tertarik untuk melakukan penelitian dengan mengajukan judul **“Analisis Sentimen Pada Ulasan Aplikasi Airbnb di Google Play Store Menggunakan Metode Naïve Bayes Classification”**. Harapannya, Analisis ini berguna untuk Airbnb maupun perusahaan pemesanan lainnya, karena motivasi konsumen dalam memberikan ulasan mengenai layanan akomodasi Airbnb ini dapat menjadikan acuan dalam perbaikan atau peningkatan dari segi fasilitas ataupun kenyamanan penggunaan aplikasi

Metode Naïve Bayes

Naïve Bayes Classifier adalah algoritma *machine learning* probabilistik yang digunakan dalam berbagai macam tugas klasifikasi. Untuk bisa memahami algoritma ini, berikut rumus umum Teorema Bayes yang menjadi dasar dari Naive Bayes berikut:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

A, B = events
 $P(A|B)$ = probability of A given B is true
 $P(B|A)$ = probability of B given A is true
 $P(A), P(B)$ = the independent probabilities of A and B

Gambar 2.2 Rumus Teorema Bayes

Teorema Bayes membantu untuk bisa mengetahui sebuah probabilitas dengan menghitung probabilitas lain yang masih berkaitan. Setelah diterapkan pada algoritma Naive Bayes, rumus ini menghasilkan suatu asumsi dasar. Dalam melihat sebuah fitur, algoritma ini selalu berasumsi bahwa fitur tersebut mandiri, setara, dan memiliki kontribusi pada hasil.

Tinjauan Studi

Penelitian ini mengumpulkan informasi dan laporan-laporan penelitian sejenis yang

dijadikan sebagai acuan dari studi literatur. Berikut Tabel 2.1 adalah beberapa penelitian sejenis yang dijadikan literatur :

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti dan Tahun	Judul	Metode	Hasil
1.	Tri Herdiawan Apandi, Castaka Agus Sugiant (2019)	Algoritma <i>Naive Bayes</i> untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP	Naïve Bayes	Algoritma Naïve Bayes lebih akurat untuk data tingkat kepuasan pelayanan e-ktip di Kecamatan Batujajar dengan akurasi sebesar 91.70% dan nilai pengujian f-measure sebesar 93,92%, sedangkan hasil dari akurasi algoritma <i>decision tree</i> 65.90%.
2.	Fithri Selva Jumeilah (2018)	Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE Dengan <i>Naïve Bayes</i>	<i>Naïve Bayes</i>	Dari hasil klasifikasi komentar pelanggan JNE diperoleh bahwa Naïve bayes mampu mengklasifikasikan data dengan baik. Hasil perhitungan ini dibuktikan dari nilai rata-rata presentase akurasi 85%, <i>precision</i> 78% dan <i>recall</i> 67%.

3.	Suwanda Aditya Saputra, Didi Rosiyadi , Windu Gata, Syepri Maulana Husain (2017)	Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes (Studi Kasus : OVO)	Naïve Bayes	Berdasarkan penelitian dari <i>cross validation</i> NB tanpa Feature Selection menghasilkan 82.30 % untuk <i>accuracy</i> dan 0.780 untuk AUC. Sedangkan hasil <i>Naive Bayes</i> dengan <i>Feature Selection</i> menghasilkan akurasi adalah 83.60 % dan 0.801 untuk AUC. Peningkatan sangat signifikan dengan penggunaan <i>Feature Selection (FS) Particle Swarm Optimization</i>
4.	Syahmia Gusriani, Kartina Diah Kusuma Wardhani, Muhammad Ihsan Zul (2017)	Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)	Naïve Bayes	Hasil analisis pengujian menunjukkan kestabilan akurasi setelah diuji dengan <i>k-Fold Cross Validation</i> dan <i>Confusion Matrix</i> dengan akurasi 93.7% dimana <i>minimum support</i> untuk <i>Frequent Itemset Mining</i> adalah 0.014

5.	Dwimarcahyani, D., Badriyah, T., & Karlita, T. (2019)	<i>Classification On Category Of Public Responses On Television Program Using Naive Bayes Method</i>	<i>Naïve Bayes</i>	Hasil dari penelitian ini adalah daftar informasi tentang tanggapan publik yang telah dikategorikan untuk memudahkan KPI atau orang-orang yang mengakses dapat menentukan program terbaik. Dari 326 jumlah data yang telah digunakan sebagai dataset dengan prosentase 80% data latih dan 20% data uji dan menghasilkan akurasi 82%.
----	---	--	--------------------	--

METODE PENELITIAN

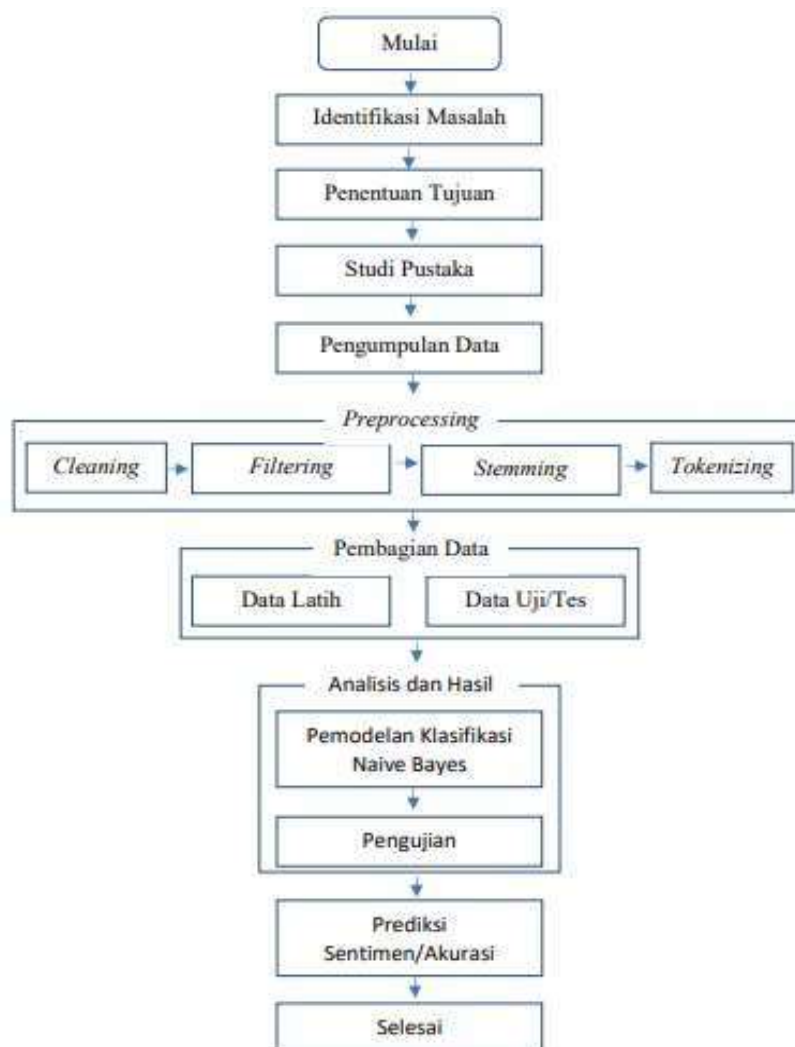
Metode Pengumpulan Data

Adapun proses pengumpulan data dalam penelitian adalah sebagai berikut :

- Studi literatur dilakukan dengan mengumpulkan dan menggali informasi akurat yang berkaitan dengan topik penelitian serta mengkaji kemajuan metodologi yang berhubungan dengan penelitian yang akan dilakukan.
- Studi Pustaka dilakukan dengan cara membaca, mengutip dan membuat catatan yang bersumber pada bahan-bahan pustaka yang mendukung dan berkaitan dengan penelitian ini khususnya dalam pengembangan Sistem Pendukung Keputusan.
- Dalam penelitian ini, data yang digunakan diperoleh menggunakan teknik web scrapping. Web Scraping merupakan metode pengambilan data dari sebuah website dan digunakan untuk analisis sentimen (Romadoni dkk., 2020).

Kerangka Pemikiran

Dalam menyelesaikan penelitian ini, penulis membuat sebuah kerangka pemikiran yang digunakan sebagai acuan dalam penelitian ini sehingga penelitian dapat dilakukan dengan baik. Permasalahan pada penelitian ini adalah belum diketahuinya metode yang tepat dengan akurasi terbaik untuk klasifikasi opini yang memiliki sentimen terhadap review pengguna aplikasi Airbnb, dengan menggunakan metode yang digunakan adalah Naïve Bayes Algorithm. Sebelum melakukan klasifikasi review yang berasal dari google play, terlebih dahulu mengenal proses pre-prosesing yang adalah cara untuk menemukan informasi yang terkandung didalam teks yang dijadikan sumber data penelitian. Berikut adalah penjelasan dari Kerangka Pemikiran :



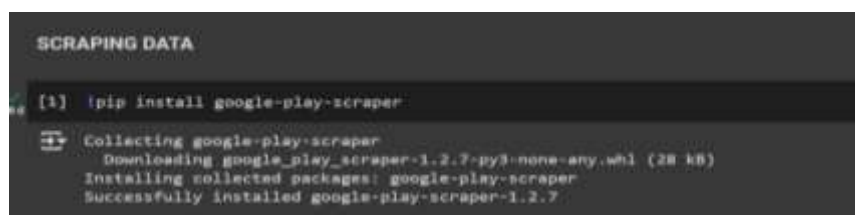
Gambar 4.3 Kerangka Pemikiran\

HASIL DAN PEMBAHASAN

• Scrapping

Berikut ini merupakan proses dalam pengambilan data dengan *scrapping* pada aplikasi *AirBnB* menggunakan *Google Colaboratory* kemudian menggunakan library “*Google Play scraper*” dalam bahasa *Python*. Data yang diambil ialah komentar dari *review* aplikasi dan penilaian *rating point*, tahapan dalam *Scrapping data* sebagai berikut :

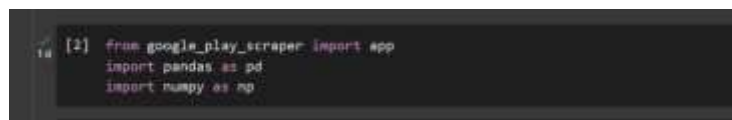
1. Menginstall *library Scrapingplaystore* pada saat pengambilan data *review* ulasan pengguna aplikasi *AirBnB*.



```
SCRAPING DATA
[1] !pip install google-play-scraper
Collecting google-play-scraper
  Downloading google_play_scraper-1.2.7-py3-none-any.whl (28 kB)
Installing collected packages: google-play-scraper
Successfully installed google-play-scraper-1.2.7
```

Gambar 5.1 Instal Scrapping Play Store

2. Mengimport modul yang akan digunakan yaitu `import pandas as pd` dan `import NumPy as np`, *Pandas (Python for Data Analysis)*



```
[2] from google_play_scraper import app
import pandas as pd
import numpy as np
```

Gambar 5.2 Kode Program Tahap Mengimport

3. Mengimportkan data dengan melakukan *Scrapping data* yaitu memasukan “ID” aplikasi, “Lang” menggunakan bahasa Indonesia, “Country” negara Indonesia, “Sort” menggunakan ulasan yang paling relevant dan “Count” 500 data ulasan yang akan di *Scrapping* pada masing-masing setiap aplikasi.



```
[3] from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.airbnb.android',

    lang='id', #untuk bahasa Indonesia
    country='id', #untuk negara Indonesia
    sort=Sort.MOST_RELEVANT, #ulasan yang paling relevant
    count=500, #data yang diambil
    filter_score_with=None #mengambil semua score atau bintang
)
```

Gambar 5.3 Kode Program Tahap Scrapping Aplikasi AirBnB

Setelah semua komentar terkumpul, selanjutnya hasil *scraping data* dapat di download pada file csv kemudian melakukan konversi ke dalam *excel* dengan format *xlsx*

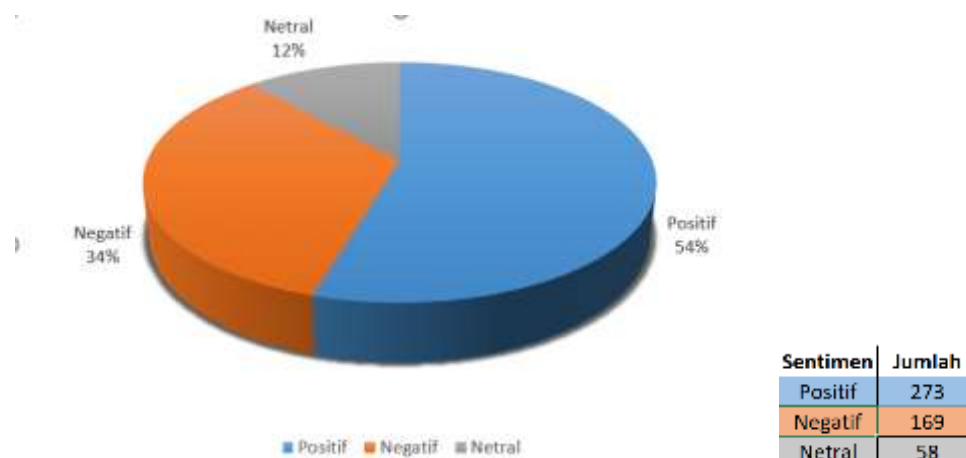
- **Pelabelan**

Data yang tersimpan hasil dari *Scraping data* selanjutnya akan dilakukan *Labelling* untuk membedakan sentimen positif dan sentimen negatif. Untuk mempermudah proses klasifikasi maka *sentimen positif* ditandai dengan label positif dengan *score* yaitu 4 dan 5, sentimen negatif ditandai dengan *score* 1 dan 2 selanjutnya memberikan nilai valuasi sentimen dengan memberikan skor *rating* yaitu *score* 1,2,3,4,5 selanjutnya menghapus bias *rating score* 3 yaitu sentimen netral pada komentar aplikasi belajar belajar pada *rating Google Play Store* syang akan mengubah analisis sentimen tersebut menjadi dua yaitu *sentimen positif* dan *analisis negatif* .

PELABELAN

```
[11] def pelabelan(score):
    if score < 3:
        return 'Negatif'
    elif score == 4 :
        return 'Positif'
    elif score == 5 :
        return 'Positif'
    my_df['Label'] = my_df ['score'].apply(pelabelan)
    my_df.head(50)
```

Gambar 5.5 Kode Program Pelabelan



Gambar 5.7 Persentase Data Setelah Labelling

Selanjutnya pada hasil proses pelabelan data sentimen aplikasi *AirBnB* menghasilkan total sebanyak 500 terdiri 273 yang bernilai komentar positif (Bintang 4-5) dan 169 komentar yang bernilai negatif (Bintang 1-2) dan 58 bernilai komentar netral (Bintang 3).

• Data Pre-Processing

a. Cleaning

Pada tahap proses *Cleaning* pada tahap ini bertujuan membersihkan ulasan atau komentar dari hal yang tidak diperlukan seperti tanda baca, *emoticon*, bahasa tidak baku dan sebagainya.



```

CLEANING

my_df['text_clean'] = my_df['content'].str.lower()
my_df['text_clean']
data_clean = clean_text(my_df, 'content', 'text_clean')
data_clean.head(10)

```

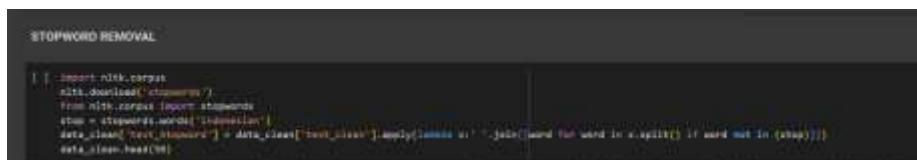
Gambar 5.8 Kode Program Cleaning

	content	score	label	text_clean
0	Aplikasi terbaru yg sangat membantu	5	Positif	aplikasi terbaru yg sangat membantu
1	bagus banget aplikasi nya ga mengecewakan	5	Positif	bagus banget aplikasi nya ga mengecewakan
2	Manfaat sangat membantu	5	Positif	manfaat sangat membantu
3	Aplikasi yang merugikan host, kalau tamu tidak...	1	Negatif	aplikasi yang merugikan host kalau tamu tidak...
4	Mudah digunakan	5	Positif	mudah digunakan
5	akhirnya Airbnb bisa pembayaran non kartu kredit...	5	Positif	akhirnya Airbnb bisa pembayaran non kartu kredit...
6	susah banget pake aplikasi ini, host daftar via...	1	Negatif	susah banget pake aplikasi ini host daftar via...
7	Sangat bermanfaat	5	Positif	sangat bermanfaat
8	Semangat pake Airbnb merasa di rugikan dan ti...	1	Negatif	semangat pake Airbnb merasa di rugikan dan ti...
9	Kali pertama menggunakan Airbnb sangat memuaskan...	5	Positif	kali pertama menggunakan Airbnb sangat memuaskan...

Gambar 5.9 Hasil Cleaning

b. Filtering / Stopword Removal

Filtering adalah tahap mengambil kata-kata penting dari hasil token dengan menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting).



```

STOPWORD REMOVAL

import nltk.corpus
nltk.download('stopwords')
from nltk.corpus import stopwords
stem = stopwords.words('indonesian')
data_clean['text_stopword'] = data_clean['text_clean'].apply(lambda x: ' '.join(word for word in x.split() if word not in (stop)))
data_clean.head(10)

```

Gambar 5.10 Kode Program Filtering/Stopword Removal

id	name	description	category	status	created_at	updated_at	deleted_at
1	Item 1	Item 1 description	Category 1	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
2	Item 2	Item 2 description	Category 2	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
3	Item 3	Item 3 description	Category 3	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
4	Item 4	Item 4 description	Category 4	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
5	Item 5	Item 5 description	Category 5	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
6	Item 6	Item 6 description	Category 6	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
7	Item 7	Item 7 description	Category 7	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
8	Item 8	Item 8 description	Category 8	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
9	Item 9	Item 9 description	Category 9	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
10	Item 10	Item 10 description	Category 10	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
11	Item 11	Item 11 description	Category 11	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
12	Item 12	Item 12 description	Category 12	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
13	Item 13	Item 13 description	Category 13	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
14	Item 14	Item 14 description	Category 14	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
15	Item 15	Item 15 description	Category 15	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
16	Item 16	Item 16 description	Category 16	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
17	Item 17	Item 17 description	Category 17	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
18	Item 18	Item 18 description	Category 18	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
19	Item 19	Item 19 description	Category 19	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	
20	Item 20	Item 20 description	Category 20	Active	2023-01-01 00:00:00	2023-01-01 00:00:00	

Gambar 5.11 Hasil Filtering/Stopword Removal

c. Stemming

Stemming adalah metode untuk menemukan kata dasar. Stemming adalah salah satu tahapan preprocessing

```

$ pip install Sastrawi
Collecting Sastrawi
  Downloading Sastrawi-1.5.1-py3.pyg-none-any.whl (285 kB)
    Installing collected packages: Sastrawi
    Successfully installed Sastrawi-1.5.1

```

Gambar 5.12 Tahapan Kode Program Instal Sastrawi

Kemudian melakukan tahapan kode pemrograman tahap *Stemming* dengan mengambil data comment dan value pada aplikasi AirBnB dengan menghapus kata imbuhan pada setiap komentar, hasil data tersebut dalam berbentuk csv

[illegible]

Gambar 5.14 Tampilan kata hasil stemming

d. Tokenizing

Tokenizing adalah proses yang memisahkan serangkaian karakter berdasarkan spasi dan memungkinkan karakter tertentu seperti tanda baca dihapus pada saat yang bersamaan.

Tokenisasi ini bertujuan agar tidak lagi memproses kalimat-kalimat pada proses berikutnya, melainkan memproses kata demi kata.



```

TOKENIZING (Proses Pemisahan Teks menjadi potongan-potongan untuk dianalisa)

[ ] In: import nltk
      nltk.download('punkt')
      from nltk.tokenize import sent_tokenize, word_tokenize
      data_clean['text_tokens'] = data_clean['text_statement'].apply(lambda x: word_tokenize(x))
      data_clean.head()
  
```

Gambar 5.15 Kode Program Tokenizing



	content	score	label	text_clean	text_statement	text_tokens
0	Aplikasi yang mengizinkan host kalau kamu tidak...	1	Negatif	aplikasi yang mengizinkan host kalau kamu tidak...	aplikasi yang mengizinkan host kalau kamu kamar...	[aplikasi, mengizinkan, host, kalau, kamu, kamar, ...]
1	Mudah digunakan	5	Positif	mudah digunakan	mudah	[mudah]
2	akutnya arsitek bisa pembayaran non kartu kredit...	5	Positif	akutnya arsitek bisa pembayaran non kartu kredit...	akutnya arsitek bisa pembayaran non kartu kredit...	[akutnya, arsitek, bisa, pembayaran, non, kartu, kredit, ...]
3	susah banget pake aplikasi ini buat daftar...	1	Negatif	susah banget pake aplikasi ini buat daftar...	susah banget pake aplikasi ini buat daftar...	[susah, banget, pake, aplikasi, ini, buat, daftar, ...]
4	Sangat bermanfaat	5	Positif	sangat bermanfaat	bermanfaat	[bermanfaat]

Gambar 5.16 Hasil Tokenizing

- **Pembobotan TF-IDF**

TF-IDF (Term Frequency Inverse Document Frequency atau Pembobotan nilai terhadap kata yang sering muncul) adalah metode pembobotan bentuk integral dari term frequency dan inverse document frequency. Penelitian ini menggunakan metode TF-IDF untuk memilih fitur sebagai hasil rangkuman dan menerapkannya pada pemilihan fitur bobot kata.

- **Klasifikasi Multinomial Naive Bayes**

Gunakan sklearn library dan algoritma multinomial *Naive Bayesian* untuk mengklasifikasikan komentar sehingga menjadi data clean, dan membaginya menjadi dua bagian, yaitu 80% dari dataset akan digunakan untuk proses pelatihan data, dan 20% sisanya akan digunakan sebagai bahan uji data.

Proses yang akan dilakukan menggunakan Multinomial *Naive Bayes* yang akan menggunakan *splitting* (80:20) yang berarti 80% data akan digunakan sebagai data latih (*training data*), dan 20% data akan digunakan sebagai *testing data*. Data yang digunakan pada aplikasi dimana awalnya adalah 500 data, dikarenakan nilai *rating* adalah bias atau *rating* dengan skor 3 (netral). Data yang digunakan pada aplikasi *AirBnB* adalah 442 data komentar karena mempunyai nilai *rating* bias sehingga 58 komentar dihilangkan dalam pengujian.

Dari hasil Pengujian menggunakan multinomial *Naïve bayes* yang dilakukan adalah sebagai berikut :

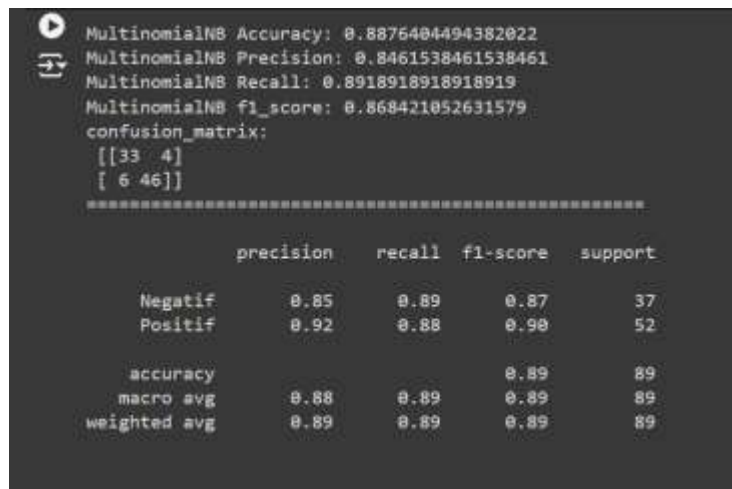
Didapatkan hasil Confusion Matrix TP sebanyak 33, FP dengan jumlah 6, FN dengan jumlah 4 dan TN dengan jumlah 46.

```
confusion_matrix:
[[33  4]
 [ 6 46]]
```

Confusion matriks :

- True Negatif (TN)*
- True Positive (TP)*
- False Negatif (FN)*
- False Positive (FP)*

- Hasil Klasifikasi / Evaluasi (Accuracy, Precision, Recall dan f1- score)***



```

MultinomialNB Accuracy: 0.8876404494382022
MultinomialNB Precision: 0.8461538461538461
MultinomialNB Recall: 0.8918918918918919
MultinomialNB f1_score: 0.868421052631579
confusion_matrix:
[[33  4]
 [ 6 46]]

```

	precision	recall	f1-score	support
Negatif	0.85	0.89	0.87	37
Positif	0.92	0.88	0.90	52
accuracy			0.89	89
macro avg	0.88	0.89	0.89	89
weighted avg	0.89	0.89	0.89	89

Gambar 5.19 Hasil Klasifikasi

Pada Nilai *Precision* (Perbandingan data antara *True Positive* “TP” dengan banyak data yang diprediksi Positif) dihasilkan menggunakan algoritma *Multinomial Naive Bayes* didapatkan sebesar 0.84615, Nilai *Recall* (Perbandingan banyak data antara *True Positive* “TP” dengan banyak data yang sebenarnya positif) yang didapatkan sebesar 0.89189 dan *F1- score* (Nilai *Harmoni* antara *Precision* dengan *Recall*) yang didapatkan sebesar 0.86842 dengan keakuratan algoritma yang digunakan sebesar 0.88764 atau sebesar 89% dapat dikatakan bahwa model klasifikasi memiliki nilai yang sangat baik.

KESIMPULAN

Berdasarkan pengujian dan analisis terhadap ulasan sentimen pada Pengguna aplikasi AirBnB dengan metode *Naive Bayes* dapat disimpulkan bahwa :

1. Dataset didapatkan dengan melakukan *scraping* di *Play Store* pada aplikasi pengguna AirBnB didapatkan 500 data ulasan berbagai pengguna yang nantinya dijadikan bahan pengujian.
2. Dengan menggunakan metode *Naive Bayes*, dari 500 dataset dilakukan penghapusan data untuk *rating point* untuk yang bernilai 3 dikarenakan nilai tersebut adalah nilai bias atau yang bernilai netral selanjutnya dataset akan dilakukan *splitting* (80:20) dari 80% dataset digunakan sebagai *data training* dan 20% dataset digunakan sebagai data *testing*.
3. Hasil pengujian menggunakan metode multinomial *Naive Bayes* yang dilakukan dengan data (20% dari dataset secara *random*) pada aplikasi AirBnB digunakan berjumlah 89 data didapatkan hasil *confusion matriks* :
Nilai True Negatif (TN) : Sebanyak 46 data
Nilai True Positive (TP) : Sebanyak 33 data
Nilai False Negatif (FN) : Sebanyak 4 data
Nilai False Positive (FP) : Sebanyak 6 data
4. Dengan menggunakan pembobotan TF-IDF didapatkan hasil :
 - a. Nilai *Precision* sebesar 0.84615
 - b. Nilai *Recall* sebesar 0.89189
 - c. Nilai *F1- score* sebesar 0.86842

Nilai tingkat keakuratan algoritma sebesar 0.88764 atau sekitar 89%. Dapat dikatakan bahwa model klasifikasi memiliki nilai yang sangat baik

DAFTAR PUSTAKA

Chohan, S., Nugroho, A., Aji, A. M. B., & Gata, W. (2020). Analisis Sentimen Pengguna Aplikasi *Duolingo* Menggunakan Metode *Naïve Bayes* dan Synthetic Minority Over Sampling Technique. *Paradigma - Jurnal Komputer dan Informatika*, 22(2), 139–144. <https://doi.org/10.31294/p.v22i2.8251>

Erfina, A., Basryah, E. S., Saepulrohman, A., & Lestari, D. (2020). Analisis Sentimen Aplikasi Pembelajaran *Online* Di *Play Store* Pada Masa Pandemi Covid-19

Menggunakan Algoritma Support Vector Machine
(Svm<http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/4094>)

Lestari, S., & Saepudin, S. (2021). Support Vector Machine: Analisis Sentimen Aplikasi Saham di *Google Play Store*. *JUSIFO (Jurnal Sistem Informasi)*, 7(2), 81–90. <https://doi.org/10.19109/jusifo.v7i2.9825>

Muslehatin, W., Ibnu, M., & Mustakim. (2017). Penerapan *Naïve Bayes* classification Untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi Uin Suska Riau | Muslehatin | Seminar Nasional Teknologi Informasi Komunikasi dan Industri. *Komunikasi dan Industri*, 250–256. <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/3276/2158>

Nurhayati, Indra, Nuralamsyah Zulkarnaim. (2021). Penerapan Algoritma *Naive Bayes* Dalam Memprediksi Penyakit Lambung. *Journal of Information System, Informatic and Computing*, 5. 524–531. <https://doi.org/10.31605/jcis.v1i1>

Petiwi, M. I., Triayudi, A., & Sholihati, I. D. (2022). Analisis Sentimen *Gofood* Berdasarkan Twitter Menggunakan Metode *Naïve Bayes* dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, 6(1), 542. <https://doi.org/10.30865/mib.v6i1.3530>

Ratnawati, F. (2018). Implementasi Algoritma *Naive Bayes* Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50. <https://doi.org/10.35314/isi.v3i1.335>

Santoso, D. P., & Wibowo, W. (2022). Analisis Sentimen Ulasan Aplikasi *Buzzbreak* Menggunakan Metode *Naïve BayesClassifier* pada Situs *Google Play Store* <https://doi.org/10.12962/j23373520.v11i2.72534>

Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government pada *Google Play* Menggunakan Algoritma *Naïve Bayes*. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 785–795. <https://doi.org/10.35957/jatisi.v9i2.1835>

Sari, A. E., Widowati, S., & Lhaksmana, K. M. (2019). Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di *Google Play Store* dengan Menggunakan Metode Information Gain dan *Naive Bayes Classifier*. *E- Proceeding of Engineering*, 6(2), 9143–9157.

Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

<https://dqlab.id/mengenal-naive-bayes-sebagai-salah-satu-algoritma-data-science>

<https://apps.apple.com/id/app/airbnb/id401626263?l=id>