

**ANALISIS SENTIMEN ULASAN APLIKASI GOJEK
DENGAN MENGGUNAKAN ALGORITMA
NAÏVE BAYES**



**UNIVERSITAS
BUDI LUHUR**

Nama Anggota Kelompok :

Febri Ardana (2212500199)
Sri Wahyuni Dian Mifella (2212501304)
Ardelia Adzra (2212500645)

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

2024

A. LATAR BELAKANG

Ulasan pengguna di Google Play merupakan penilaian subjektif yang diberikan oleh individu setelah menggunakan suatu aplikasi atau game. Para pengguna memberikan tinjauan ini berdasarkan pengalaman mereka dengan berbagai aspek, mulai dari kinerja aplikasi, fitur-fitur yang disediakan, hingga respons dari pengembang. Ulasan ini tidak hanya memberikan umpan balik langsung kepada pengembang, tetapi juga memengaruhi persepsi dan keputusan pengguna potensial dalam memilih untuk mengunduh atau tidak suatu aplikasi. Dengan demikian, ulasan pengguna di Google Play memiliki peran penting dalam membentuk reputasi dan keberhasilan suatu aplikasi di platform tersebut. Sentimen analisis adalah studi komputasional dari opini-opini orang, sentiment melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif.

Seiring bertambahnya sumber opini yang tidak terbatas, penentuan sentimen secara manual menjadi tidak efisien. Oleh karena itu, metode *machine learning*, seperti Naïve Bayes classifier digunakan untuk mengklasifikasikan polaritas opini dari banyak sumber data. Metode ini memiliki kecepatan dan akurasi tinggi, cocok untuk data besar dan beragam. Kelebihannya meliputi kesederhanaan, kecepatan, dan akurasi yang tinggi. Sehingga dalam menyelesaikan permasalahan tersebut dilakukan "Analisis Sentimen Ulasan Aplikasi Gojek di Google Playstore dengan Menggunakan Algoritma Naive Bayes", Penelitian ini bertujuan untuk melakukan analisis opini pengguna terkait dengan aplikasi Gojek berbasis sentimen analisis.

1. RUMUSAN MASALAH

Berdasarkan latar belakang diatas telah didapatkan beberapa masalah yang dapat dirumuskan sebagai berikut :

- a) Bagaimana cara mendapatkan data ulasan pengguna dari google play store?
- b) Bagaimana cara menganalisis sentimen pengguna yang menggunakan aplikasi Gojek?

2. METODOLOGI PENELITIAN

Setelah mengetahui permasalahan yang ingin dianalisis terdapat beberapa metode untuk memperoleh data atau informasi dalam menyelesaikan permasalahan. Metode yang dapat dilakukan antara lain :

- a) Sumber Data

Google playstore menjadi sumber data untuk melakukan analisis sentimen dikarenakan permasalahan yang ingin dianalisis adalah opini atau ulasan pengguna selama menjadi pengguna aplikasi tersebut.

- b) Pengumpulan Data

Data yang dikumpulkan berdasarkan ulasan pengguna terkait opini yang paling relevan atau paling banyak disetujui pengguna lain dengan cara melakukan *web scraping* dengan google colab.

- c) Pre-processing

Tools yang digunakan sebelum data diolah menggunakan metode Naïve Bayes adalah Google Colab. Data mentah diproses melalui text preprocessing kemudian diberi label.

3. TUJUAN PENELITIAN

Dengan adanya beberapa permasalahan yang terbentuk, dapat disimpulkan bahwa tujuan diadakannya penelitian adalah untuk mengetahui bagaimana hasil dari analisis sentimen berdasarkan ulasan pengguna yang paling relevan dengan pengguna lain sehingga dapat diketahui performa dari aplikasi Gojek selama ini.

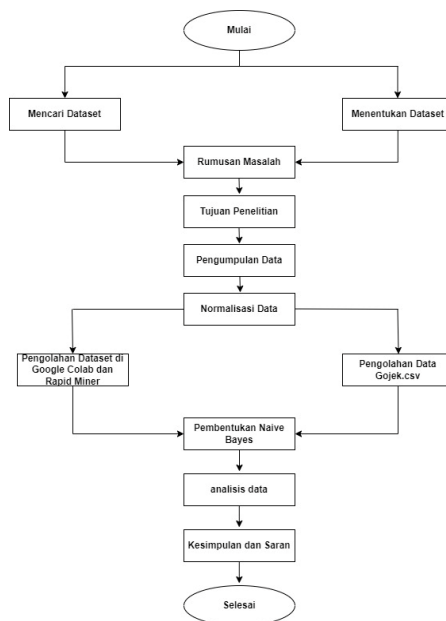
B. TAHAPAN PENELITIAN

Link scraping :

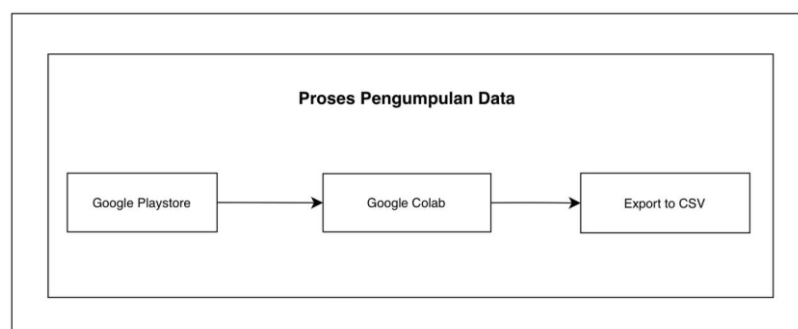
https://colab.research.google.com/drive/1z0bkETDw_yKFo6IBS7bV7piK-jc3PPXH?usp=sharing

a) Pengumpulan Data

Dalam penelitian ini data yang diambil adalah ulasan dan rating dari pengguna aplikasi Gojek melalui *web scraping*. Data yang didapatkan dari web scraping adalah 1000 data opini dalam bentuk komentar. Setelah mendapatkan data tersebut langkah selanjutnya yaitu melakukan text processing agar dapat dengan mudah memberi label positif dan negatif berdasarkan rating ulasan pengguna.



Gambar 1 : Flowchart Pemrosesan data



Gambar 2 : Proses Pengumpulan data

Proses pengumpulan data ini dilakukan dari google playstore, kemudian dilakukan scrapping menggunakan program python yang ada di google colab kemudian hasil dari data scrapping tersebut disimpan dengan format CSV.

```
Scraping Kelompok Gojek
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
#scrape jumlah ulasan yang diinginkan
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.gojek.app',
    lang='id', #disini kita mau men scrape data ulasan aplikasi shopee yang berada di google play store
    country='id', #kita setting bahasa nya menjadi bahasa Indonesia
    sort=Sort.MOST_RELEVANT, # # kemudian kita gunakan most_relevant untuk mendapatkan ulasan yang paling relevant
    count=1000, # disini jumlah ulasan yang mau kita ambil ada seribu
    filter_score_with=None # # kemudian di filter_score kita gunakan None untuk mengambil semua score atau rating bintang 1 sampai 5
)
```

Gambar 2 : Proses Web Scraping menggunakan Google Colab

b) Pembersihan data

Pada tahapan ini data kotor yang didapatkan dari proses scraping harus dibersihkan. Data kotor menyimpan banyak sekali kolom yang tidak diperlukan untuk kebutuhan analisa, tetapi untuk *missing values* telah dilakukan pengecekan dan hasilnya tidak ada *missing values* pada dataset ini.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	40126724-6e07-420a-bd82-ba899a29d078	Zaky Naufal Anshari	lh.googleusercontent.com/a-/ALV-U...	Aplikasi sudah bagus, sayangnya tidak diberika...	3	5	4.87.3	2024-05-29 02:31:46	Hai Kak Zaky, mohon maaf atas ketidaknyamanann...	2024-05-29 03:49:56	4.87.3
1	6630bbcc-1a46-4c91-9b5f-368fcfc6294	Noviano	lh.googleusercontent.com/a/ACg8oc...	Jujur agak kecewa Sama maps nya Go-Jek, rumah d...	2	4	4.87.3	2024-05-29 09:59:36	Hai, Kak Noviano. Pastikan kamu sudah memberik...	2024-05-29 15:07:15	4.87.3
2	52574be6-0921-442a-87a5-d6c0ba69d25c	Kak Rurr_	lh.googleusercontent.com/a-/ALV-U...	Pelayanan nggk jelas, sekarang mau batalkan pe...	1	1	4.85.3	2024-05-29 01:23:18	Hai Kak Rur, mohon maaf atas ketidaknyamanann...	2024-05-29 07:17:13	4.85.3
3	806232ce-416b-41c6-8665-45ea480a4d8f	Ristia Ayu Nandira	lh.googleusercontent.com/a/ACg8oc...	Sebelum langganan GOJEKPLUS sudah jarang dapat...	1	75	4.87.2	2024-05-13 04:46:30	Hai Kak Ristia, terkait laporanmu sudah ada tl...	2024-05-14 16:27:46	4.87.2
4	07417ab6-c81e-4fa6-9755-dcc8b14224dd	Aydalah Dalahay	lh.googleusercontent.com/a-/ALV-U...	Entah lah males sama kesel banget sama gojek k...	1	1	4.88.1	2024-05-30 05:47:44	Hai Kak Aydalah, jika lupa PIN, klik 'Lupa PIN...	2024-05-30 09:02:33	4.88.1



Gambar 3 : Hasil data set dari melakukan web scraping dengan menggunakan Google Colab dan pengecekan missing values

Gambar diatas adalah hasil data scrapping ulasan di playstore pada aplikasi sebesar 1000 data. Pada data tersebut sudah tidak ada missing values sehingga data dapat langsung diolah sesuai kebutuhan. Untuk

mengetahui ulasan positif dan negatif, diperlukan filtering data pada kolom score agar hanya data dengan score 1-2 dan 4-5 saja yang ditampilkan.

```
[ ] df_busu[['userName', 'score', 'at', 'content']].head() #dari scrapping tsb didapatkan banyak sekali kolom, kemudian kolom" tsb kita filter
#sehingga didapatkan kolom username, score, at dan content
```

	userName	score	at	content
0	Zaky Naufal Anshari	3	2024-05-29 02:31:46	Aplikasi sudah bagus, sayangnya tidak diberika...
1	Noviano	2	2024-05-29 09:59:36	Jujur agak kecewa Sama maps nya Go-Jek, rumah d...
2	Kak Rurr_	1	2024-05-29 01:23:18	Pelayanan nggk jelas, sekarang mau batalkan pe...
3	Ristia Ayu Nandira	1	2024-05-13 04:46:30	Sebelum langganan GOJEKPLUS sudah jarang dapat...
4	Aydaiah Dalahay	1	2024-05-30 05:47:44	Entah lah males sama kesel banget sama gojek k...

Gambar 4 : Proses dari data cleaning yaitu filter kolom yang ingin digunakan

Gambar diatas adalah data yang sudah dilakukan *cleaning* tetapi masih terdapat score netral atau score 3 pada data, sehingga harus dilakukan filtering dan pelabelan untuk score netral.

PELABELAN

```
def pelabelan(score):
    if score < 3:
        return 'Negatif'
    elif score == 1 :
        return 'Negatif'
    elif score == 4 :
        return 'Positif'
    elif score == 5 :
        return 'Positif'
my_df['Label'] = my_df ['score'].apply(pelabelan)
my_df.head(50)
```

	content	score	Label
590	Pengelolaan tambah gk jelas, pesen gocar susah...	1	Negatif
888	Kapok beli kuota internet di fitur GoPulsa, sa...	1	Negatif
34	aplikasi ppk, gua top up lewat bca mbanking ke...	1	Negatif
645	Kenapa jarak resto di aplikasi beda dengan jar...	4	Positif
891	Jgn langganan goplus Discount plus... Rugi cum...	2	Negatif
903	Sangat membantu di saat lg tidak ada kendaraan...	5	Positif
914	Iklan yang otomatis di layar sangat mengganggu	1	Negatif
910	Bagus, bagus sekali, sangat bagus. Banyakin pr...	5	Positif

Gambar 5 : Proses dari data cleaning yaitu memberi pelabelan pada score

Gambar diatas adalah program untuk melakukan pelabelan data yang akan menghilangkan data score netral atau score 3, jadi hanya data dengan range score 1-2 dan 4-5 saja yang ditampilkan.

Row No.	Label	content	score
1	Positif	Membantu sekali	5
2	Negatif	Saya lg kesel ikan anda muncul trs di hp tolong hentikan	1
3	Positif	Sangat membantu	5
4	Positif	Jelek	4
5	Positif	Ok	5
6	Positif	Oke punya, semoga berkah	5
7	Positif	Bagus puas	5
8	Positif	Recommended	5
9	Negatif	ongkir makin hari makin mahal dan ongkos ojeknya juga makin mahal	2
10	Positif	Bagus	5
11	Positif	Banyak promonya mantap	5
12	Negatif	akun terblokir tidak terbuka buka padahal saya sudah lapor tapi tidak ada tindakan tolong lah di tindak lanjut	1
13	Negatif	Untuk titik lokasi melenceng, padahal sudah disesuaikan manual, tapi baik lagi ke titik awal yang melenceng	1
14	Positif	Mantap gojek	5
15	Negatif	cs nya makan gaji buta. kompln ga drespon!!!	1
16	Positif	Joossss	5
17	Positif	Gohek matap	5

Row No.	Label	content	score
19	Positif	Ok banget gan	5
20	Positif	Apknya bagus dan sangat berguna	5
21	Negatif	Heloooo... Makin sini makin mehoooll aja ya biaya penanganan dan pengiriman wkwkwkwk	2
22	Positif	Sy senang menggunakan tranfortasi gojek apalagi ada diskon Waah mantab deh pokonya	5
23	Negatif	Gojek kenapa? pesen goride aja 30 menit gak daper driver, padahal banyak driver di sekitar. Parties mau bangkrut.	1
24	Positif	Bagus	4
25	Negatif	Semakin jelek sangat jelek, skrg gojek kalo perjalanan dekat ngak ada yg mau terima. Kecewa jelek jlek jelek. Udh jelek pilih2 lagi. Ngak bak	1
26	Positif	Mantab	5
27	Negatif	Jelek banget pesen pake express nunggu nya malah lebih lama dari pada hemat terus gunanya ada express, hemat, reguler apa coba g	1
28	Positif	Sangat" membantu ...	5
29	Negatif	Sumpah ga profesional banget ... tiba tiba di blokir ... padahal pernah telat bayar dan aplikasi pun terus kasih kenaikan limit tapi saat ini tiba	1
30	Positif	Membantu	5
31	Positif	Ko enggak pernah dapet vocer lagi ya, apa karena aku pernah komplain gara-gara makannya enggak enak 8'ye"	5
32	Positif	Sangat nyaman dan puas	5
33	Positif	baik, ga ngebut"	5
34	Positif	Respon cpt kmna pun jd mudah	5
35	Positif	Sangat membantu	5

Gambar 6 : Hasil dataset yang telah melalui proses data cleaning (data bersih)

Gambar diatas adalah hasil data yang sudah melalui tahapan *cleaning* yang akan di proses kedalam rapid miner untuk diolah menggunakan algoritma naive bayes, data tersebut harus mempunyai label untuk patokan ke setiap datanya yang akan di olah, sehingga kolom label dijadikan label agar dapat diketahui total ulasan positif dan negatif.

c) Pre-processing data

Dalam tahapan ini dilakukan beberapa metode untuk membuat data lebih terstruktur dengan melakukan beberapa tahap seperti stopwords removal, tokenizing, stemming.

Stopword Removal

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopwords dalam bahasa Indonesia adalah "yang", "dan", "di", "dari", dll. Makna di balik penggunaan stopwords yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya.

```
import nltk.corpus
nltk.download('stopwords')
from nltk.corpus import stopwords
stop = 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(50)
```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.

	userName	score	at	content	Label	text_clean	text_Stopword
0	Fasya Suroso	1	2024-06-17 13:01:33	Gopaylater saya diblokir padahal saya bayar te...	Negatif	gopaylater saya diblokir padahal saya bayar te...	gopaylater diblokir bayar kalo gak lg mending ...
1	lusyawatty joice	1	2024-06-17 11:48:15	Parah banget Nic Go-Jek, aplikasinya eror , ma...	Negatif	parah banget nic gojek aplikasinya eror masa ...	parah banget nic gojek aplikasinya eror ngak p...
2	HAMS	3	2024-06-17 08:23:37	Tingkatkan terus akselerasi, akurasi, keamanan...	Netral	tingkatkan terus akselerasi akurasi keamanan d...	tingkatkan akselerasi akurasi keamanan fitur a...
3	Dam Jor	1	2024-06-17 07:16:33	Aplikasi mulai ga jelas,udah pesan makanan ham...	Negatif	aplikasi mulai ga jelasudah pesan makanan hamp...	aplikasi ga jelasudah pesan makanan jam ga dri...
4	Busyro Hady	5	2024-06-17 06:10:55	Saya penguna baru Go-Jek, tapi saya bisa paham...	Positif	saya penguna baru gojek tapi saya bisa paham b...	penguna gojek paham membantu berbelanja di te...

Executing (46s) <cell line: 24> > write() > schedule() > send()

Gambar 7 : Dilakukan proses stopwords removal untuk menghilangkan kata-kata yang tidak memiliki makna atau dianggap tidak penting (tahap pre-processing)

TOKENIZING

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian di analisa. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Didalam NLP, token diartikan sebagai "kata" meskipun tokenize juga dapat dilakukan pada paragraf maupun kalimat

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

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.

	userName	score	at	content	Label	text_clean	text_StopWord	text_tokens
0	Fasya Suroso	1	2024-06-17 13:01:33	Gopaylater saya diblokir padahal saya bayar te...	Negatif	gopaylater saya diblokir padahal saya bayar te...	gopaylater diblokir bayar kalo gak lg mending ...	[gopaylater, diblokir, bayar, kalo, gak, lg, m...
1	lusyawatty joice	1	2024-06-17 11:48:15	Parah banget Nic Go-Jek, aplikasinya eror , ma...	Negatif	parah banget nic gojek aplikasinya eror masa ...	parah banget nic gojek aplikasinya eror ngak p...	[parah, banget, nic, gojek, aplikasinya, eror,...
2	HAMS	3	2024-06-17 08:23:37	Tingkatkan terus akselerasi, akurasi, keamanan...	Netral	tingkatkan terus akselerasi akurasi keamanan d...	tingkatkan akselerasi akurasi keamanan fitur a...	[tingkatkan, akselerasi, akurasi, keamanan, fi...
3	Dam Jor	1	2024-06-17 07:16:33	Aplikasi mulai ga jelas,udah pesan makanan ham...	Negatif	aplikasi mulai ga jelasudah pesan makanan hamp...	aplikasi ga jelasudah pesan makanan jam ga dri...	[aplikasi, ga, jelasudah, pesan, makanan, jam,...
4	Busyro Hady	5	2024-06-17 06:10:55	Saya penguna baru Go-Jek, tapi saya bisa paham...	Positif	saya penguna baru gojek tapi saya bisa paham b...	penguna gojek paham membantu berbelanja di te...	[penguna, gojek, paham, membantu, berbelanja, ...

Gambar 8 : Dilakukan proses tokenizing untuk pemisahan teks menjadi potongan kata-kata (tahap pre-processing)


```

Scraping Kelompok Gojek
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
RAM
Disk
+ Gemini

[46] # apply stemmed term to dataframe
def get_stemmed_term(document):
    return [term_dict[term] for term in document]

#script ini bisa dipisah dari eksekusinya setelah pembacaan term selesai
data_clean['text_stemindo'] = data_clean['text_tokens'].apply(lambda x: ' '.join(get_stemmed_term(x)))
data_clean.head(20)

4254
-----
1 : gopaylater : gopaylater
2 : diblokir : blokir
3 : bayar : bayar
4 : kalo : kalo
5 : gak : gak
6 : lg : lg
7 : mending : mending
8 : uninstall : uninstall
9 : aja : aja
10 : deh : deh
11 : males : males
12 : pake : pake
13 : nya : nya
14 : gapernah : gapernah
15 : telat : telat
16 : blokir : blokir
17 : aplikasi : aplikasi
18 : aneh : aneh
19 : parah : parah
20 : banget : banget
21 : nic : nic
22 : gojek : gojek
23 : aplikasinya : aplikasi

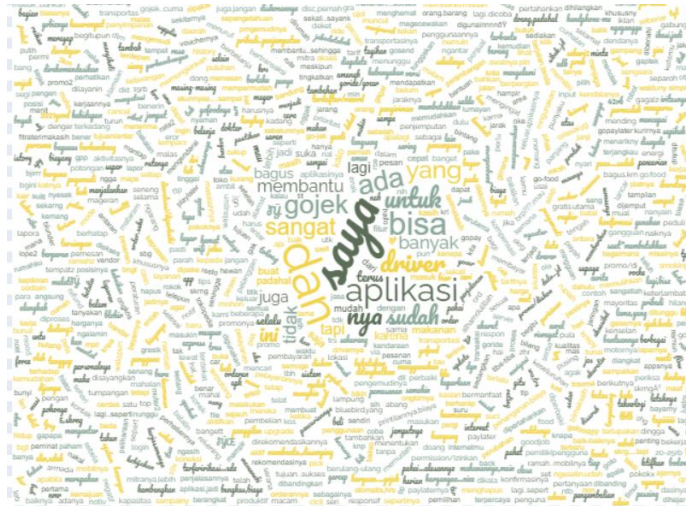
```

Gambar 9 : Dilakukan proses stemming (pre-processing)



Gambar 10 : Wordclouds untuk komentar dengan ulasan negatif

Gambar diatas adalah hasil dari wordclouds yang menampilkan kata-kata paling banyak dituliskan dalam ulasan dengan range score 1-2.



Gambar 11 : Wordclouds untuk komentar dengan ulasan positif

Gambar diatas adalah hasil dari wordclouds yang menampilkan kata-kata paling banyak dituliskan dalam ulasan dengan range score 4-5.

```

MultinomialNB Accuracy: 0.74
MultinomialNB Precision: 0.513789128314268
MultinomialNB Recall: 0.4724747474747475
MultinomialNB f1_score: 0.4666069263612882
confusion_matrix:
[[131  0  1]
 [ 25  0  3]
 [ 23  0 17]]

```

	precision	recall	f1-score	support
Negatif	0.73	0.99	0.84	132
Netral	0.00	0.00	0.00	28
Positif	0.81	0.42	0.56	40
accuracy			0.74	200
macro avg	0.51	0.47	0.47	200
weighted avg	0.64	0.74	0.67	200

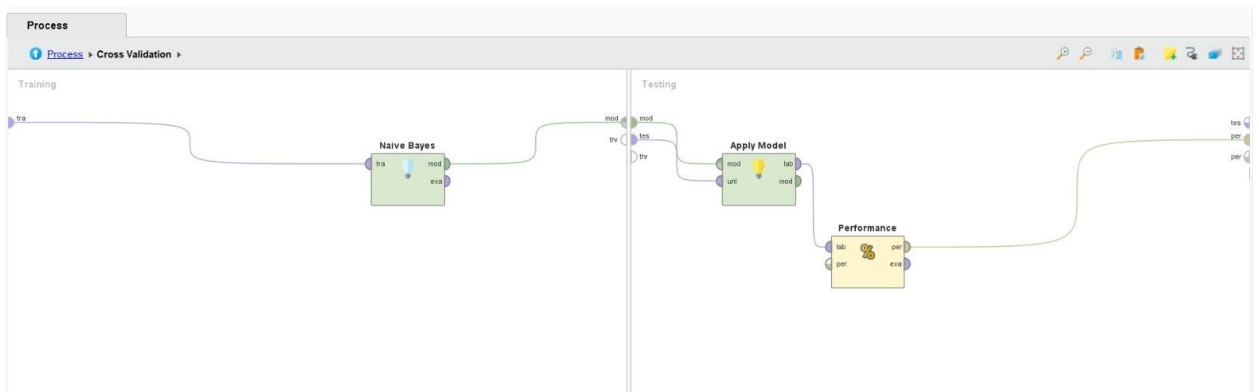
Gambar 12 : Hasil dari confusion matrix

C. VISUALISASI DATA



Gambar 13 : Proses melakukan cross validation melalui aplikasi Rapid Miner

Gambar diatas adalah processing yaitu dengan menggunakan read csv untuk membuka file csv yang berisi dataset bersih, kemudian dihubungkan dengan *cross validation*.



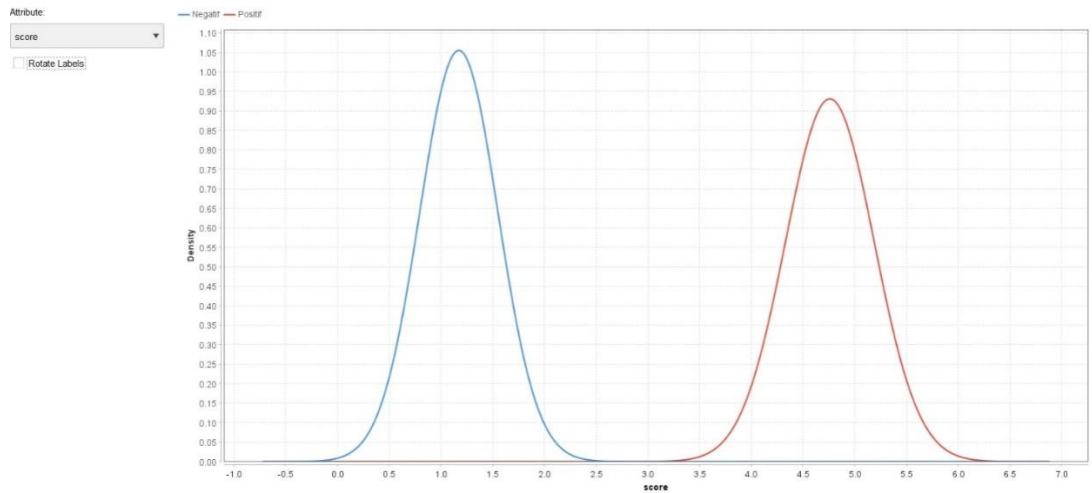
Gambar 14 : Proses melakukan cross validation melalui algoritma Naive Bayes

Gambar diatas adalah proses yang terdiri dari 2 tahap ,yaitu tahap training dan testing. Pada tahap training, data dilatih menggunakan metode Naive Bayes. Kemudian pada tahap testing, data diproses melalui Apply Model dan Performance Test. Dari proses cross validation yaitu training dan testing, Metode Naïve Bayes dari nilai akurasinya yaitu sebagai berikut

The screenshot shows the 'PerformanceVector (Performance)' window in Rapid Miner. It displays a table of performance metrics for the Naive Bayes model. The table includes columns for predicted and actual classes, and a row for class precision. The accuracy is 100.00%.

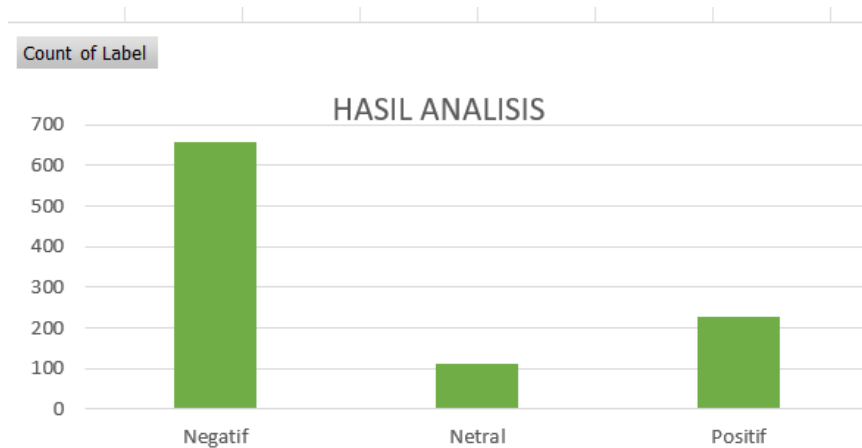
	true Negatif	true Netral	true Positif	class precision
pred. Negatif	656	0	0	100.00%
pred. Netral	0	113	0	100.00%
pred. Positif	0	0	227	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 15 : Hasil dari perfomance yang menampilkan perbandingan positif dan negatif



Gambar 16 : Visualisasi data menggunakan grafik dari aplikasi Rapid Miner

Gambar diatas merupakan grafik dari hasil analisis sentimen ulasan positif dan negatif pada aplikasi Gojek di google play store. Terdapat 658 ulasan negatif dan 229 ulasan positif. Sehingga dapat disimpulkan bahwa ulasan pengguna paling relevan selama ini menunjukkan lebih banyak sentimen negatif.



Gambar 17 : Visualisasi data menggunakan excel.

D. KESIMPULAN

Berdasarkan analisa sentimen ulasan aplikasi Gojek pada Google Play Store, dapat dilihat bahwa hasilnya cenderung menunjukkan tingkat kepuasan yang rendah dari pengguna. Dari total ulasan yang dianalisis, terdapat 658 ulasan negatif dan hanya 229 ulasan positif.

Hal ini menunjukkan adanya banyak pengguna yang mengalami kekecewaan atau masalah dalam penggunaan aplikasi Gojek. Kemungkinan masalah yang dihadapi bisa bervariasi, mulai dari performa aplikasi yang kurang memuaskan, masalah teknis, kinerja layanan pelanggan yang buruk, hingga keluhan tentang fitur-fitur yang kurang sesuai dengan harapan pengguna.

Jumlah ulasan negatif yang signifikan dapat menggambarkan bahwa ada aspek-aspek tertentu dalam pengalaman pengguna dengan aplikasi Gojek yang perlu diperbaiki atau ditingkatkan. Analisa ini dapat menjadi titik awal bagi pengembang untuk melakukan evaluasi mendalam terhadap umpan balik pengguna dan mengimplementasikan perbaikan yang dibutuhkan untuk meningkatkan kepuasan dan pengalaman pengguna secara keseluruhan untuk *developer* aplikasi Gojek.