

Implementasi Text Mining untuk Analisis Sentimen Masyarakat terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store

¹Kamelia Cindy Astuti, ²Andri Firmansyah, ³Agus Riyadi
^{1,2,3}Universitas Pelita Bangsa
Bekasi, Indonesia

¹kcindyastuti@gmail.com, ²andrifirmansyah@pelitabangsa.ac.id,
³agusriyadi@pelitabangsa.ac.id

*Penulis Korespondensi

Diajukan : 24/01/2024

Diterima : 26/01/2024

Dipublikasi : 27/01/2024

ABSTRAK

Aplikasi Digital Korlantas Polri merupakan sebuah aplikasi resmi yang digagas oleh Korlantas Polri Indonesia yang bertujuan untuk memudahkan masyarakat Indonesia untuk mendapatkan layanan di Korlantas, salah satunya adalah untuk pembuatan dan perpanjangan SIM (Surat Izin Mengemudi). Pada aplikasi Digital Korlantas Polri terdapat berbagai macam ulasan atau *review* para pengguna mengenai kinerja aplikasi tersebut yang disertai dengan rating (*score*) yang beragam dari ulasan yang terdapat pada Google Play Store. Banyaknya komentar masyarakat dalam menilai aplikasi tersebut menjadi acuan penulis dalam melakukan analisis sentimen. Analisis sentimen ini dilakukan untuk mengetahui sentimen positif maupun sentimen negatif. Pada penelitian ini dilakukan analisis sentimen dengan mengimplementasikan text mining dan pengujiannya menggunakan metode klasifikasi dari Algoritma *Naïve Bayes*. Dataset yang digunakan pada penelitian ini menggunakan hasil *Scraping* dengan menggunakan Google Colab yang berjumlah 5000 data, yang diambil dengan teknik *Web Scraping*, dimana data yang diambil adalah data yang paling relevan. Dari data tersebut, dilakukan pembagian data, yaitu data *Training* 80% dan data *Testing* 20%. Data yang akan dilakukan dalam pengujian adalah data *Testing* dimana data tersebut telah dilakukan *preprocessing* data dan diperoleh sebanyak 890 data. Hasil pengujian pada penelitian ini menggunakan perhitungan *Confusion Matrix* dan diperoleh hasil nilai *accuracy* sebesar 88%, *precision* 88%, *recall* 91%, dan *F1-Score* sebesar 90%.

Kata Kunci: Analisis Sentimen, Digital Korlantas Polri, *Naïve Bayes*, Klasifikasi, Text Mining

I. PENDAHULUAN

Pada perkembangan teknologi yang semakin pesat dalam era digital saat ini, semakin banyak aplikasi-aplikasi berbasis mobile yang memberikan kemudahan bagi masyarakat dalam berbagai hal pada kehidupan sehari-hari. Google Play Store adalah sebuah platform dari Google yang sangat populer untuk mengunduh aplikasi dimana di dalamnya terdapat fitur pengguna dapat memberikan ulasan terhadap aplikasi yang telah digunakannya. Berdasarkan portal data statistik, pada periode Desember tahun 2023 telah tersedia sekitar 2.430.000 aplikasi mobile pada Google Play Store (Ceci, 2024). Hal yang menarik dari aplikasi Google Play adalah adanya sebuah fitur yang dapat berisi ulasan (*review*) dari para penggunanya. Hal ini dapat dimanfaatkan untuk melihat ulasan atau *review* dari pengguna aplikasi tertentu, salah satunya adalah aplikasi Digital Korlantas Polri. Digital Korlantas Polri merupakan sebuah aplikasi resmi yang dirilis oleh Korlantas Polri Indonesia yang bertujuan untuk memudahkan masyarakat Indonesia untuk

mendapatkan berbagai layanan di Korlantas (*Digital Korlantas POLRI*, 2022), seperti proses pembuatan dan perpanjangan SIM (Surat Izin Mengemudi) secara online. Pada Google Play Store, aplikasi Digital Korlantas Polri terdapat lebih dari 85.000 (delapan puluh lima ribu) ulasan yang berisi komentar tekstual dari para penggunanya. Komentar tekstual tersebut menuai berbagai macam ulasan dari para pengguna aplikasi Digital Korlantas Polri seperti komentar positif, negatif, maupun netral. Ulasan menjadi sumber informasi penting bagi pengguna lain sebelum memutuskan untuk menggunakan aplikasi tersebut. Ribuan ulasan dari pengguna juga sangat dibutuhkan guna meningkatkan pengembangan terhadap kinerja dan kualitas aplikasi. Banyaknya ribuan ulasan masyarakat terhadap aplikasi Digital Korlantas Polri, sehingga menyebabkan sulitnya menentukan kecenderungan sentimen ulasan masyarakat secara manual yang menjadi acuan bagi penulis untuk melakukan analisis sentimen. Dengan adanya permasalahan tersebut, maka dibutuhkan sebuah analisis sentimen untuk dapat mengolah sejumlah komentar dengan menerapkan text mining. Text mining sendiri adalah sebuah teknik untuk mengubah teks tak terstruktur menjadi bentuk yang lebih terstruktur yang mencakup beberapa tahap teknik sehingga dapat dianalisis dengan lebih mudah. Penerapan dari text mining dapat digunakan untuk proses analisis sentimen. Analisis sentimen adalah proses untuk mengidentifikasi dan mengelompokkan sentimen yang terkandung dalam teks, seperti ulasan, apakah pendapat yang diungkapkan dalam teks tersebut bersifat positif, negatif, atau netral. Analisis sentimen menjadi proses yang sangat penting dalam memahami isi data dengan tujuan mengekstrak informasi atau menggali pengetahuan dari sumber data tekstual yang ada dalam jumlah besar (Fathonah & Herliana, 2021). Metode yang digunakan pada penelitian ini adalah metode *Naive Bayes*, algoritma tersebut dianggap cocok digunakan untuk menganalisis sentimen yang ada pada dataset. *Naive Bayes* pun memiliki performa yang baik dalam melakukan klasifikasi tanpa adanya penambahan metode lain (Fathonah & Herliana, 2021). Melalui metode klasifikasi tersebut, maka akan dilakukan perhitungan *Confusion Matrix* dan diperoleh hasil akurasi terhadap pengujian algoritma *Naive Bayes* terhadap penelitian ini.

II. STUDI LITERATUR

Penelitian Terdahulu

Berdasarkan penelitian terdahulu dengan judul “Analisis Sentimen Aplikasi *E-Government* Pada Google Play Menggunakan Algoritma *Naive Bayes*”. Pada penelitian ini, dataset yang digunakan sebanyak 642 data ulasan pada aplikasi Sentuh Tanahku, dan pembagian datanya menjadi 577 data latih 65 data uji, kemudian dihasilkan akurasi sebesar 89%, presisi sebesar 83%, dan *recall* sebesar 87%, dengan hasil tersebut dapat disimpulkan bahwa klasifikasi menggunakan algoritma *Naive Bayes* mampu memprediksi dengan cukup baik (Tanggraeni & Sitokdana, 2022).

Penelitian selanjutnya adalah mengenai Analisis Sentimen Ulasan Aplikasi *My Pertamina* Pada Google Play Store Menggunakan Algoritma NBC (*Naive Bayes Classifier*), Hasil analisis sentimen aplikasi *My Pertamina* menggunakan metode algoritma *Naive Bayes* menghasilkan *accuracy* 87 %, *precision* 86%, *recall* 90%, dan *F1-Score* 87%. Dari data yang berdasarkan label positif dan negatif ulasan aplikasi, didapatkan hasil analisis sentimen aplikasi *My Pertamina* cenderung bersifat negatif (Maulana et al., 2023).

Penelitian terdahulu lainnya yaitu mengenai Analisis Sentimen Pada Ulasan Aplikasi *Mobile Banking* Menggunakan Metode *Naive Bayes* Dengan Kamus *Inset*, mayoritas ulasan pada aplikasi tersebut yaitu ulasan positif dimana hasil dari klasifikasi tersebut menghasilkan nilai akurasi 93,1%, *precision* 90,4%, *recall* 100%, dan *f-measure* 95%. Data latihnya adalah data ulasan yang telah diberi label secara otomatis menggunakan kamus *Inset* yang telah melalui penyesuaian kata dan bobot (Nadira et al., 2023).

Text Mining

Text Mining adalah suatu proses penambangan intisari dari suatu dokumen data berupa teks yang bentuknya lebih tidak teratur yang dapat dibuat pola untuk menghasilkan sebuah informasi yang berguna. Proses penambangan data dapat bersumber melalui *website*, yang datanya dapat berupa data dokumen berbentuk teks dari *website* yang di dalamnya berisi ulasan, opini, kritik dan feedback. Text mining merupakan salah satu ragam dari data mining yang bertujuan untuk

mengambil format yang dapat diolah dari sekumpulan data-data yang berupa text dalam jumlah besar (Maulana et al., 2023).

Text Preprocessing

Tahap pertama dalam melakukan text mining yaitu text *preprocessing*. Tahapan *preprocessing* merupakan tugas penting yang harus dilakukan sebelum data digunakan dalam pembuatan model. Persiapan text dan penghapusan merupakan bagian dari pemrosesan data dan proses analisis yang melibatkan informasi dari kumpulan data yang digunakan dan merencanakan *preprocessing* yang dilakukan (Hermawan & Jowensen, 2023). Secara umum proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut :

1. *Case Folding*

Case Folding adalah proses penyamaan case dalam sebuah dokumen. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *Case Folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar. Bentuk standar tersebut berupa huruf kecil (Kartikasari et al., 2020).

2. *Stopword Removal (Filtering)*

Tahap filtrasi adalah tahap mengambil kata-kata penting dari hasil *token*. Algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata yang penting) dapat digunakan pada tahap ini. *Stopword* adalah kata-kata yang tidak deskriptif dan bukan merupakan kata penting dari suatu dokumen sehingga dapat dibuang. Contoh *Stopword* adalah “yang”, “dan”, “di”, “dari”, dan lain sebagainya (Saputra et al., 2022).

3. *Tokenizing*

Tokenizing adalah proses memecah kalimat menjadi kata-kata (*token*) yang dilakukan untuk menjadikan sebuah kalimat menjadi lebih bermakna.

4. *Stemming*

Stemming merupakan proses mengembalikan suatu kata kepada bentuk dasarnya. Jika dalam Bahasa Indonesia, proses stemming dapat digunakan untuk menghilangkan imbuhan yang menyertai kata dasarnya. *Stemming* Bahasa Indonesia yang digunakan sebagai acuan adalah *Library Sastrawi* yang disiapkan untuk Bahasa Pemrograman *Python* (Saputra et al., 2022).

Analisis Sentimen

Analisis sentimen merupakan sebuah teknik *Natural Language Processing* (NLP) yang dapat digunakan untuk mengatasi permasalahan tersebut. Analisis sentimen biasanya digunakan pada data berbentuk teks untuk menentukan apakah suatu data bersifat positif, negatif, atau netral. Tujuan dari dilakukannya analisis sentimen adalah untuk membantu sebuah bisnis atau perusahaan dalam memonitor sentimen dari *customer feedback* terhadap brand itu sendiri ataupun suatu produk, dan juga membantu perusahaan untuk memahami kebutuhan pelanggan dan mengambil keputusan bisnis (Nadira et al., 2023).

Analisis sentimen adalah proses mengidentifikasi sentimen dalam teks dengan mengolah sebuah data teks untuk memahami sentimen yang terkandung dalam sentimen tersebut. Analisis sentimen dapat digunakan untuk menganalisis opini yang terdapat dalam ulasan Google Play Store yang ditulis oleh pengguna melalui proses penggalian informasi (Friska Aditia Indriyani et al., 2023).

Naïve Bayes

Naïve Bayes adalah algoritma yang mengandalkan kemungkinan atau probabilitas dengan pengklasifikasi statistik kelas data, klasifikasi ini dikelompokkan ke kelas tertentu sesuai teori probabilitas. Algoritma *Naïve Bayes* dapat menarik kesimpulan berdasarkan klasifikasi data *Training* yang disimpan sebelumnya. Meskipun sifat independen antara kata (istilah) atau parameter dalam dokumen tidak sepenuhnya terpenuhi dengan sempurna, tapi kinerja *Naïve Bayes* untuk klasifikasi relatif dapat diandalkan dan bahkan lebih unggul dalam hal realistik,

kecepatan tinggi dan akurasi yang baik. Metode klasifikasi dapat menemukan pola suatu model dengan membedakan suatu konsep atau kelas yang tidak diketahui namanya dan termasuk ke dalam metode data mining. *Naïve Bayes* ini memprediksi probabilitas peluang di masa mendatang bersumber pada pengalaman yang sudah ada sebelumnya (Maulana et al., 2023).

Klasifikasi *Naïve Bayes* adalah salah satu metode klasifikasi yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*, klasifikasi ini didasarkan pada teorema *Bayes* yang memiliki akurasi dan kecepatan yang tinggi saat diimplementasikan kedalam *database* dengan dataset yang besar (Fitri, 2020). Algoritma ini termasuk dalam *supervised learning* karena membutuhkan data *training* sebelum proses klasifikasi. Secara umum rumus dasar persamaan teorema Bayes adalah sebagai berikut (Suhendra et al., 2021):

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

Keterangan:

- X = Data dengan kelas yang belum diketahui
- H = Hipotesis X merupakan suatu kelas spesifik
- $P(H|X)$ = Probabilitas H berdasarkan kondisi X
- $P(H)$ = Probabilitas pada hipotesis H (prior)
- $P(X|H)$ = Probabilitas X berdasarkan hipotesis H
- $P(X)$ = Probabilitas X (data sampel yang diamati)

Evaluasi

Tahap evaluasi adalah tahap yang dilakukan untuk melihat seberapa baik performansi algoritma klasifikasi yang digunakan dalam penelitian. Tahapan evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil dibuat pada langkah sebelumnya (Herlinawati et al., 2020). Tolok ukur yang digunakan untuk mengukur performansi tersebut adalah *accuracy*, *precision*, *recall*, dan *F1-Score* (Tanggraeni & Sitokdana, 2022). Untuk menghitung berdasarkan tolok ukur tersebut dibutuhkan metode *Confusion Matrix*. *Confusion Matrix* merupakan suatu matriks pengukuran performa yang berisi nilai perkiraan benar dan nilai perkiraan salah dimana keluarannya dapat berupa dua kelas atau lebih (Maulana et al., 2023). Berikut adalah model dasar dari tabel *Confusion Matrix* (berukuran 2 x2) :

Tabel 1. *Confusion Matrix* ukuran 2x2

Actual Class	Prediction Class	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Performa yang diperoleh dari *Confusion Matrix* adalah seperti akurasi, presisi, *recall*, dan *F1-score*. Berikut adalah persamaan untuk menghitung performa tersebut (Tanggraeni & Sitokdana, 2022):

a. *Accuracy*

Accuracy merupakan mengukur sejauh mana model berhasil untuk klasifikasi dokumen-dokumen dengan benar pada seluruh dataset. Berikut adalah rumus dari perhitungan akurasi :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

b. *Precision*

Presisi digunakan untuk sejauh mana dokumen-dokumen yang diprediksi sebagai suatu kelas benar-benar termasuk ke dalam kelas tersebut. Berikut adalah rumusnya :

$$Precision = \frac{TP}{TP+FP}$$

c. *Recall*

Recall merupakan mengukur sejauh mana dokumen-dokumen dari suatu kelas berhasil diidentifikasi oleh model. Berikut adalah rumusnya :

$$Recall = \frac{TP}{TP+FN}$$

d. *F1-Score*

F1-Score merupakan kombinasi dari *Precision* dan *Recall* untuk memberikan gambaran keseluruhan kinerja model. Berikut rumusnya :

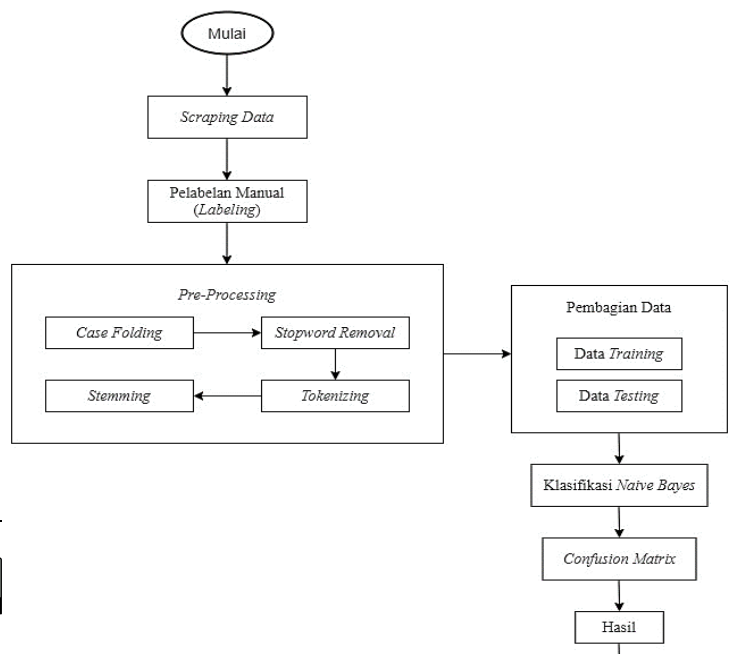
$$F1-Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

WordCloud

WordCloud salah satu dari metode text mining yang menampilkan kata-kata populer terkait dengan kata kunci internet dan data teks. Metode ini digunakan untuk memvisualisasikan data teks secara visual. *WordCloud* populer dalam text mining karena mudah dipahami. Dengan menggunakan *WordCloud*, gambaran frekuensi kata-kata dapat ditampilkan dalam bentuk yang menarik namun tetap informatif. Ukuran gambar teks dalam *WordCloud* menyesuaikan dengan frekuensi data, semakin banyak frekuensi kata digunakan, maka semakin besar pula ukuran kata tersebut ditampilkan dalam *WordCloud* (Pradana, 2020).

III. METODE

Pada penelitian ini, metode yang digunakan adalah menggunakan algoritma *Naïve Bayes*. Berikut adalah gambar metode pada penelitian ini:



Gambar 1 Metode Penelitian

Berikut adalah uraian dari gambar di atas :

Scraping Data

Tahap pertama dalam pengujian adalah proses *Web Scraping* data menggunakan Google Colab. *Web Scraping* merupakan proses pengambilan data dari sebuah web dengan cara ekstrasi. Data yang dimaksud pada penelitian ini adalah data ulasan (*review*) pada aplikasi Digital Korlantas Polri. *Keyword* yang digunakan adalah url dari *website* aplikasi Digital Korlantas Polri pada Google play store. Pada penelitian ini, data ulasan yang akan diambil adalah ulasan yang paling relevan (*Most Relevant*) dan jumlah data yang akan diambil adalah sebanyak 5000 data ulasan. Berikut adalah proses *web Scraping* dengan Google Colab :

```

from google_play_scraper import Sort, reviews
result, continuation_token = reviews(
    'id.qoin.korlantas.user',
    lang = 'id',
    country = 'id',
    sort = Sort.MOST_RELEVANT,
    count = 5000,
    filter_score_with = None
)
    
```

Gambar 2 Proses *Web Scraping*

Pelabelan (*Labeling*)

Setelah dataset terkumpul, pada penelitian ini proses selanjutnya adalah pemberian *Labeling* (pelabelan) dimana proses *Labeling* dilakukan secara manual yaitu dengan memberikan label positif dan label negatif. Label positif dilihat dari pemberian *score* di atas 3 sampai 5, sedangkan untuk label negatif dari *score* 1 sampai dengan 3. Berikut adalah hasil pemberian label pada dataset:

```

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)
    
```

	content	score	Label
0	Aplikasi ga bisa masuk	3	None
1	Udh dari tgl 09 Desember 2023 ngajuin, tp blm ...	1	Negatif
2	Kecewa kenapa digitalisasi SIM selalu gagal. S...	1	Negatif
3	Masa udah tes sgala macem... udah bayar biaya ...	1	Negatif
4	Perpanjangan SIM sangat dimudahkan lewat layan...	5	Positif
5	Alhamdulillah sangat sangat membantu, mudah, c...	5	Positif
6	Setelah dibayar statusnya stuck 40% gak dipros...	2	Negatif
7	Bukanya mempermudah malah jadi panjag banyakan ...	1	Negatif
8	Polri lebih maju dan semoga kedepan qda terobo...	5	Positif

Gambar 3 Pemberian Label pada Dataset

Preprocessing Data

Text preprocessing adalah suatu proses untuk menyeleksi data text agar menjadi lebih terstruktur lagi dengan melalui serangkaian tahapan. Pada penelitian ini tahapan-tahapannya yaitu, Case Folding, Stopword Removal, Tokenizing, dan Stemming.

Case Folding

Case Folding merupakan sebuah proses untuk mengubah huruf kapital menjadi huruf kecil (lowercase). Pada penelitian ini, data ulasan tidak hanya berubah menjadi huruf kecil saja, melainkan karakter seperti tanda baca, angka, dan karakter lainnya telah dibersihkan pada tahap ini yaitu yang disebut dengan proses Cleaning. Berikut adalah proses Case Folding :

	content	score	Label	text_clean
1	Udh dari tgl 09 Desember 2023 ngajuin, tp blm ...	1	Negatif	udh dari tgl desember ngajuin tp blm di pros...
2	Kecewa kenapa digitalisasi SIM selalu gagal. S...	1	Negatif	kecewa kenapa digitalisasi sim selalu gagal su...
3	Masa udah tes sgala macem... udah bayar biaya ...	1	Negatif	masa udah tes sgala macem udah bayar biaya kes...
4	Perpanjangan SIM sangat dimudahkan lewat layan...	5	Positif	perpanjangan sim sangat dimudahkan lewat layan...
5	Alhamdulillah sangat sangat membantu, mudah, c...	5	Positif	alhamdulillah sangat sangat membantu mudah cep...
6	Setelah dibayar statusnya stuck 40% gak dipros...	2	Negatif	setelah dibayar statusnya stuck gak diproses ...
7	Bukanya mempermudah malah jadi panjang banyakan ...	1	Negatif	bukanya mempermudah malah jadi panjang banyakan ...
8	Polri lebih maju dan semoga kedepan qda terobo...	5	Positif	polri lebih maju dan semoga kedepan qda terobo...
9	Saya coba pake apl ini, sudah selesai semua sa...	1	Negatif	saya coba pake apl ini sudah selesai semua sam...
11	Kenapa kok mau daftar buat sim gak bisa dan ad...	2	Negatif	kenapa kok mau daftar buat sim gak bisa dan ad...
13	Perpanjang SIM C ditolak. Sudah bayar dan test...	1	Negatif	perpanjang sim c ditolak sudah bayar dan test ...
14	Proses cepat dan mudah, pelayanan baik	5	Positif	proses cepat dan mudah pelayanan baik
15	Udah sampai tahap masukan rekening pengembalia...	1	Negatif	udah sampai tahap masukan rekening pengembalia...

Gambar 3 Proses Case Folding

Stopword Removal

Stopword Removal proses membuang kata-kata kurang penting atau menyimpang dari kosa kata yang tidak memiliki arti. Berikut adalah hasil dari proses Stopword Removal:

text_clean	text_StopWord
udh dari tgl desember ngajuin tp blm di proses msh di tanya sampe kali alasannya selalu karna sedang banyak yg melakukan pengajuan tp masa iya udh sebulan lebih msh ga di proses jg ga ada kejelasan diterima apa engga nya berkas yg udh di buat	udh tgl desember ngajuin tp blm proses msh sampe kali alasannya karna yg pengajuan tp iya udh sebulan msh ga proses jg ga kejelasan diterima engga nya berkas yg udh
kecewa kenapa digitalisasi sim selalu gagal sudah di lakukan secara berkala berbulan tetep aja gagal serius gak sih app ini bisa buat digitalisasi sim	kecewa digitalisasi sim gagal lakukan berkala berbulan tetep aja gagal serius gak sih app digitalisasi sim
masa udah tes sgala macem udah bayar biaya kesehatan maju kena mundur kena cancel gabisa diterusin gabisa yang niat apa kalo bikin aplikasi	udah tes sgala macem udah bayar biaya kesehatan maju kena mundur kena cancel gabisa diterusin gabisa niat kalo bikin aplikasi
perpanjangan sim sangat dimudahkan lewat layanan online ini sim a c selalu lewat online terima kasih korlantas polri the best pokoknya	perpanjangan sim dimudahkan layanan online sim a c online terima kasih korlantas polri the best pokoknya
alhamdulillah sangat sangat membantu mudah cepat dan transparasi setelah dibayar statusnya stuck gak diproses dah berhari lokasi ditlantas daan mogot padahal sbntr lagi expired bingung klo stuck smp expired trus ditolak dibatalin pun gak iso	alhamdulillah membantu mudah cepat transparasi dibayar statusnya stuck gak diproses dah berhari lokasi ditlantas daan mogot sbntr expired bingung klo stuck smp expired trus ditolak dibatalin gak iso
bukanya mempermudah malah jadi panjang banyakan dokumen yang di minta tidak mebanbu harusnya klo digital cukup photo diri ktp sim gitu aja ko repot klo harus ini itu dokumen laien sma ja dateng ke tempat jaa langsgn perpajang tete susah	bukanya mempermudah panjang banyakan dokumen mebanbu klo digital photo ktp sim gitu aja ko repot klo dokumen laien sma ja dateng jaa langsgn perpajang tete susah
polri lebih maju dan semoga kedepan qda terobosan yg lebih baikmakasih digital korlantas polri	polri maju semoga kedepan qda terobosan yg baikmakasih digital korlantas polri
saya coba pake apl ini sudah selesai semua sampe pembayaran tpi beberapa hari kemudian ditolak tanpa pemberitahuan di apl dan akhirnya suruh bikin baru gmana ini pak	coba pake apl selesai sampe pembayaran tpi ditolak pemberitahuan apl suruh bikin gmana

Gambar 4 Proses Stopword Removal

Tokenizing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai *token* untuk kemudian dianalisa. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai *token*.

text_StopWord	text_tokens
udh tgl desember ngajuin tp blm proses msh sam...	[udh, tgl, desember, ngajuin, tp, blm, proses,...
kecewa digitalisasi sim gagal lakukan berkala ...	[kecewa, digitalisasi, sim, gagal, lakukan, be...
udah tes sgala macem udah bayar biaya kesehata...	[udah, tes, sgala, macem, udah, bayar, biaya, ...
perpanjangan sim dimudahkan layanan online sim...	[perpanjangan, sim, dimudahkan, layanan, onlin...
alhamdulillah membantu mudah cepat transparasi	[alhamdulillah, membantu, mudah, cepat, transp...

Gambar 5 Proses *Tokenizing*

Stemming

Stemming merupakan proses untuk mengubah semua kata menjadi kata dasar. *Stemming* antara satu bahasa dengan bahasa yang lain tentunya berbeda. Untuk isi ulasan dengan berbahasa Indonesia tergolong lebih rumit atau kompleks karena terdapat berbagai variasi imbuhan yang harus dibuang untuk mendapatkan kata dasar.

text_clean	text_StopWord	text_tokens	text_steamindo
udh dari tgl desember ngajuin tp blm di pros...	udh tgl desember ngajuin tp blm proses msh sam...	[udh, tgl, desember, ngajuin, tp, blm, proses,...	udh tgl desember ngajuin tp blm proses msh sam...
kecewa kenapa digitalisasi sim selalu gagal su...	kecewa digitalisasi sim gagal lakukan berkala ...	[kecewa, digitalisasi, sim, gagal, lakukan, be...	kecewa digitalisasi sim gagal laku kala bulan ...
masa udah tes sgala macem udah bayar biaya kes...	udah tes sgala macem udah bayar biaya kesehata...	[udah, tes, sgala, macem, udah, bayar, biaya, ...	udah tes sgala macem udah bayar biaya sehat ma...
perpanjangan sim sangat dimudahkan lewat layan...	perpanjangan sim dimudahkan layanan online sim...	[perpanjangan, sim, dimudahkan, layanan, onlin...	panjang sim mudah layan online sim a c online ...
alhamdulillah sangat sangat membantu mudah membantu mudah cep...	alhamdulillah membantu mudah cepat transparasi	[alhamdulillah, membantu, mudah, cepat, transp...	alhamdulillah bantu mudah cepat transparasi
setelah dibayar statusnya stuck gak diproses ...	dibayar statusnya stuck gak diproses dah berha...	[dibayar, statusnya, stuck, gak, diproses, dah...	bayar status stuck gak proses dah hari lokasi ...

Gambar 6 Proses *Stemming*

Pembagian Data (Data Training dan Data Testing)

Setelah melakukan tahap *Preprocessing*, selanjutnya adalah melakukan pengujian dengan metode yang diusulkan yaitu *Naive Bayes*. Untuk dapat melakukan perhitungan Algoritma *Naive Bayes* dengan. Dari data yang sudah diberi label sentimen, data perlu dibagi menjadi dua bagian yaitu data *Training* dan data *Testing*. Pada penelitian ini penulis membagi data *Training* dan data *Testing* dengan perbandingan rasio 80% : 20%.

Evaluasi

Untuk mengetahui hasil evaluasi, pada penelitian ini menggunakan metode *Naive Bayes* dengan nantinya akan dilakukan perhitungan dengan *Confusion Matrix* untuk mengetahui hasil akurasi, *precision*, dan *recall*.

IV. HASIL DAN PEMBAHASAN

Berdasarkan hasil *Scraping* data yang telah dilakukan, diperoleh dataset sebanyak 5000 data,

kemudian setelah melalui hasil *Preprocessing* Data, jumlah data menjadi 4449 data. Dari jumlah dataset akhir tersebut dihasilkan pembagian data, yaitu data *Training* sebanyak 3559 data, dan untuk data *Testing* sebanyak 890 data. Berdasarkan pembagian data tersebut, dihasilkan jumlah sentimen positif dan sentimen negatif. Untuk jelasnya dapat dilihat pada tabel berikut :

Data Hasil Scraping	Data Akhir	Data Training	Data Testing	Positif	Negatif
5000	4449	3559	890	371	519

Gambar 7 Perolehan Dataset

Setelah dilakukan pembagian data tersebut, tahap selanjutnya adalah pengujian dataset menggunakan klasifikasi *Naïve Bayes*, dan akan dilakukan evaluasi dengan perhitungan *Confusion Matrix*. Pada penelitian ini dilakukan satu kali pengujian untuk menghitung tingkat *accuracy*, *precision*, *recall*, dan *F1-Score* dengan rasio data *Training* 80% dan data *Testing* 20%. Berikut adalah hasil evaluasi dari perhitungan *Confusion Matrix* :

```
MultinomialNB Accuracy: 0.8842696629213483
MultinomialNB Precision: 0.8866171003717472
MultinomialNB Recall: 0.9190751445086706
MultinomialNB f1_score: 0.902554399243141
confusion_matrix:
[[477 42]
 [ 61 310]]
-----
              precision    recall  f1-score   support

   Negatif      0.89      0.92      0.90       519
   Positif      0.88      0.84      0.86       371

 accuracy              0.88              0.88       890
 macro avg              0.88              0.88       890
 weighted avg           0.88              0.88       890
```

Gambar 8 Hasil *Confusion Matrix*

Dengan proses perhitungan manual sebagai berikut :

1. *Accuracy*

Accuracy merupakan menggambarkan seberapa besar akurat suatu model dapat mengklasifikasikan dengan benar. Berikut adalah nilai prediksi akurasi pada pengujian data *Testing* dengan cara perhitungan secara manual :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$= \frac{477+310}{477+310+61+42} \times 100\%$$

$$= \frac{787}{890} \times 100\%$$

$$Accuracy = 88\%$$

2. *Precision*

Precision mengevaluasi kemampuan tingkat ketelitian data antara data aktual dengan hasil prediksi dalam model untuk menemukan peringkat yang paling relevan. Berikut adalah hasil perhitungan nilai *precision* secara manual :

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$= \frac{477}{477+61} \times 100\%$$

$$= \frac{477}{538} \times 100\%$$

Score sebesar 90%. Selain itu berdasarkan data *Testing* sebanyak 890 data, dihasilkan opini berbentuk positif sebanyak 317 dan opini berbentuk negatif sebanyak 519. Sehingga dapat disimpulkan bahwa kecenderungan opini masyarakat terhadap penggunaan aplikasi Digital Korlantas Polri adalah cenderung bersifat negatif.

VI. REFERENSI

- Ceci, L. (2024). *Jumlah aplikasi yang tersedia di Google Play Store dari Desember 2009 hingga Desember 2023*. <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/>
- Digital Korlantas POLRI*. (2022). <https://www.digitalkorlantas.id/>
- Fathonah, F., & Herliana, A. (2021). Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes. *Jurnal Sains Dan Informatika*, 7(2), 155–164. <https://doi.org/10.34128/jsi.v7i2.331>
- Fitri, E. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naive Bayes*, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, 18(1), 71. <https://doi.org/10.26623/transformatika.v18i1.2317>
- Friska Aditia Indriyani, Ahmad Fauzi, & Sutan Faisal. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine. *TEKNOSAINS : Jurnal Sains, Teknologi Dan Informatika*, 10(2), 176–184. <https://doi.org/10.37373/tekno.v10i2.419>
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293. <https://doi.org/10.24114/cess.v5i2.18186>
- Hermawan, A., & Jowensen, I. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *Jurnal Sains Dan Teknologi*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- Kartikasari, T. S., Setiawan, H., & Lucky Tirma Irawan, P. (2020). Implementasi Text Mining Untuk Analisis Opini Publik Terhadap Calon Presiden. *Jurnal Simantec*, 7(1), 39–47. <https://doi.org/10.21107/simantec.v7i1.6528>
- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC. *Jurnal Teknologi Terpadu*, 9(1), 42–48. <https://doi.org/10.54914/jtt.v9i1.609>
- Nadira, A., Setiawan, N. Y., & Purnomo, W. (2023). Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Naïve Bayes Dengan Kamus Inset. *Indexia*, 5(01), 35. <https://doi.org/10.30587/indexia.v5i01.5138>
- Pradana, M. G. (2020). Penggunaan Fitur Wordcloud dan Document Term Matrix dalam Text Mining. *Jurnal Ilmiah Informatika*, 8(1), 38–43.
- Saputra, I., Djatna, T., Siregar, R. R. A., Kristiyanti, D. A., Yani, H. R., & Riyadi, A. A. (2022). Text Mining of PeduliLindungi Application Reviews on Google Play Store. *Faktor Exacta*, 15(2), 1979–276.
- Suhendra, M., Swastika, W., & Subianto, M. (2021). Analisis Sentimen Pada Ulasan Aplikasi Video Conference Menggunakan *Naive Bayes*. *Sainsbertek Jurnal Ilmiah Sains & Teknologi*, 2(1), 1–9. <https://doi.org/10.33479/sb.v2i1.145>

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>