

Analisis Ulasan Pengguna Aplikasi Seabank Dengan *Support Vector Machine* Dan *Naïve Bayes*

Cindy Nada Adela^{1*}, Sri Karnila², Sutedi³, Melda Agarina⁴

^{1, 3, 4}Ilmu Komputer, Sistem Informasi, IIB Darmajaya, Lampung, Indonesia

²Ilmu Komputer, Sains Data, IIB Darmajaya, Lampung, Indonesia

Email: ^{1*}cindynadaadela27@gmail.com, ²srikarnila_dj@darmajaya.ac.id, ³utedi@darmajaya.ac.id,

⁴meldaagarina.mti@gmail.com

Abstrak - Aplikasi perbankan digital seperti Seabank, telah menjadi bagian dari kebutuhan kehidupan saat ini. Diluncurkan pada Februari 2021 oleh PT Bank Seabank Indonesia, dan meraih popularitas yang signifikan. Penggunaan aplikasi ini menghasilkan sejumlah besar ulasan yang mencerminkan sentimen pengguna terkait dengan kinerja, layanan dan keamanan aplikasi. Memahami sentimen positif, netral, dan negatif dari ulasan pengguna. Seabank memungkinkan untuk mendapatkan wawasan tentang tingkat kepuasan pelanggan, kelemahan aplikasi, dan area perbaikan yang diperlukan. Untuk itu penting dilakukan analisis data dan klasifikasi ulasan pengguna. Penelitian ini bertujuan mengetahui kinerja dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) dan Naïve Bayes, dalam mengklasifikasikan sentimen ulasan pengguna. Pengujian dilakukan dengan menggunakan 3789 data, pembagian rasio 80% training dan 20% testing. Hasil pelabelan menunjukkan adanya 438 sentimen positif, 1379 sentimen netral, dan 77 sentimen negatif. Sedangkan hasil pelabelan oleh ahli bahasa menunjukkan jumlah 1100 sentimen positif, 308 sentimen netral, dan 486 sentimen negatif. Hasil pengujian akurasi menggunakan confusion matrix menunjukkan bahwa algoritma Support Vector Machine memiliki nilai akurasi tertinggi sebesar 63%, sedangkan algoritma Gaussian Naïve Bayes memiliki nilai terendah sebesar 30%. Dengan demikian, dapat disimpulkan bahwa model SVM lebih efektif dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Seabank daripada model Naïve Bayes.

Kata kunci: Analisis Sentimen, ulasan, *Support Vector Machine*, *Naïve Bayes*

Abstract- Digital banking applications, such as Seabank, have become part of today's needs. Launched in February 2021 by PT Bank Seabank Indonesia and has gained significant popularity. Usage of this application generates a large number of reviews reflecting user sentiment regarding the application's performance, service, and security. Understand positive, neutral, and negative sentiments from user reviews. Seabank makes it possible to gain insight into customer satisfaction levels, application weaknesses, and areas of improvement needed. For this reason, it is important to analyze the data and classify user reviews. This research aims to determine the performance of two classification algorithms, namely Support Vector Machine (SVM) and Naïve Bayes, in classifying user review sentiment. Testing was carried out using 3789 data points, a ratio of 80% training and 20% testing. The labeling results showed that there were 438 positive sentiments, 1379 neutral sentiments, and 77 negative sentiments. Meanwhile, the results of labeling by linguists showed a total of 1100 positive sentiments, 308 neutral sentiments, and 486 negative sentiments. The results of accuracy testing using the confusion matrix show that the Support Vector Machine algorithm has the highest accuracy value of 63%, while the Gaussian Naïve Bayes algorithm has the lowest value of 30%. Thus, it can be concluded that the SVM model is more effective in classifying the sentiment of Seabank application user reviews than the Naïve Bayes model.

Keyword: Sentiment Analysis, Reviews, Support Vector Machine, Naïve Bayes

1. PENDAHULUAN

Industri perbankan merupakan industri keuangan yang fokus pada kegiatan penyimpanan, penyaluran, dan pengelolaan modal. Perbankan digital menjadi arus utama ketika bank dan perusahaan teknologi berkolaborasi untuk menyediakan layanan keuangan yang lebih sederhana, lebih cepat, dan lebih terjangkau [1]. Seperti yang diketahui, Sea Group merupakan holding company di industri yang berfokus pada bidang teknologi dan IOT (Internet of Thing), contohnya Garena yang berfokus pada bidang *developing game*, Shopee dibidang *e-commerce* hingga yang terbaru Seabank dibidang perbankan digital [2]. Aplikasi perbankan digital telah menjadi alat penting dalam kehidupan sehari-hari. Salah satu aplikasi perbankan digital yang populer saat ini adalah Seabank. Seabank dapat diakses melalui aplikasi smartphone menawarkan berbagai layanan dan produk keuangan. Bank memiliki fokus yang kuat pada inovasi, berinvestasi dalam teknologi untuk meningkatkan layanannya dan memberikan kenyamanan lebih kepada pelanggan [3].

Seabank telah mengumpulkan banyak ulasan pengguna berdasarkan sentimen pengguna terkait layanan kinerja dan keamanan aplikasi seabank. Dengan memahami sentimen positif, netral dan negatif yang diungkapkan oleh pengguna aplikasi seabank dapat memperoleh wawasan tentang kepuasan pelanggan, kelemahan aplikasi, dan area perbaikan yang mungkin diperlukan. Permasalahan yang muncul adalah bagaimana mengolah banyak ulasan dari aplikasi seabank dalam jumlah besar dan melihat tren sentimen yang muncul. Untuk mengatasi permasalahan tersebut penting bantuan teknologi untuk menghasilkan model klasifikasi data ulasan. Metode *Naïve Bayes* dan *Support Vector Machine* yang akan di gunakan untuk analisis sentimen ulasan pengguna aplikasi

seabank, karena kesederhanaan, kecepatan dalam pelatihan dan prediksinya, serta kemampuannya memberikan interpretasi probabilistik dari hasil klasifikasi dan kemampuannya mengatasi data berdimensi tinggi. Dengan mempertimbangkan kelebihan dari masing-masing metode, SVM dan Naive Bayes metode yang paling efektif dan efisien untuk analisis sentimen ulasan pengguna aplikasi SeaBank. Penggunaan kedua metode ini diharapkan dapat memberikan hasil yang akurat dan dapat digunakan untuk meningkatkan kualitas layanan aplikasi.

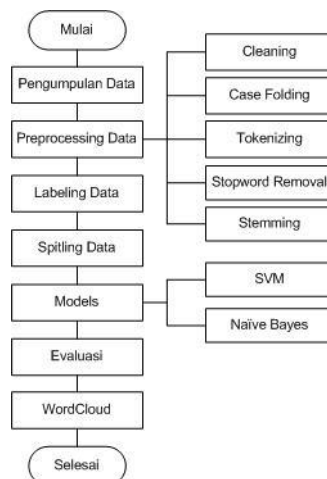
Pengolahan data ulasan dikerjakan melalui analisis sentiment, dan dasar analisis sentimen mulai dari mengumpulkan data atau teks yang ada menjadi sebuah kalimat atau dokumen, yang kemudian mengkategorikan kalimat atau dokumen tersebut menjadi bentuk positif, negatif atau netral [4]. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian ataupun aplikasi mengenai analisis sentimen berkembang pesat, bahkan di Amerika kurang lebih 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen [5].

Penelitian lain mengenai perbandingan kinerja Perbandingan Metode *Naive Bayes* dan *Support Vector Machine* pada Analisis Sentimen Twitter. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* memiliki akurasi tertinggi, yaitu 73,65% [6]. Selanjutnya penelitian Perbandingan Metode Klasifikasi Support Vector Machine dan *Naive Bayes* untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store, hasil penelitian menunjukkan bahwa algoritma SVM memiliki akurasi yang lebih tinggi dibandingkan dengan *Naive Bayes* classifier yaitu 81,46% [7].

Penelitian ini bertujuan menganalisis sentimen terhadap ulasan pengguna aplikasi Seabank. Dalam analisis ini terdapat tiga kelas sentimen yang diidentifikasi, yaitu positif, netral, dan negatif. hasilnya perbandingan kinerja model *Support Vector Machine* (SVM) dan *Naive Bayes*. Dalam perbandingan ini, dilakukan evaluasi dan perhitungan akurasi dari kedua metode untuk menentukan metode mana yang memberikan hasil yang lebih akurat dalam mengklasifikasikan sentimen ulasan pengguna aplikasi seabank. Hasil analisis dengan *Support Vector Machine* (SVM) memiliki nilai akurasi yang paling tinggi yaitu mencapai 63.2%. Sementara, nilai akurasi tertinggi dari *Naive Bayes* terdapat pada model Bernoulli *Naive Bayes*, dengan nilai akurasi 61.1%. Namun akurasi terendah terdapat pada model Gaussian *Naive Bayes*, dengan nilai akurasi sebesar 30.3%, sedangkan akurasi Multinomial *Naive Bayes* adalah 0.583, ini menunjukkan hasil kinerja SVM lebih baik dibandingkan dengan *Naive Bayes* pada penelitian ini.

2. METODE PENELITIAN

Pada bagian ini akan dijelaskan mengenai perancangan sistem dalam penelitian ini. Metode penelitian ini dilakukan melalui beberapa tahap. Secara garis besar, alur penelitian dapat dijelaskan melalui diagram pada Gambar 1.



Gambar 1. Alur penelitian analisis sentimen ulasan pengguna aplikasi seabank

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah ulasan pengguna aplikasi seabank di google play store rentang waktu dari 2022-2023. Pengumpulan data dilakukan melalui teknik web scrapping dengan menggunakan metode perhitungan menggunakan Google Colab dan diprogram dengan bahasa Python, data yang didapatkan berjumlah 3804 ulasan kemudian disimpan kedalam format.csv.

2.2 Preprocessing Data

Pada tahap ini dilakukan proses pembersihan pada data yang masih kotor. Proses pembersihan yang dilakukan pada penelitian ini antara lain [8].

- *Cleaning* adalah proses pembersihan data, Pada tahap ini penghapusan simbol, angka, dan tanda baca.
- *Case Folding* adalah proses mengubah huruf besar menjadi huruf kecil atau standar [9].
- *Tokenizing* adalah tahapan yang bertujuan untuk membagi teks dari ulasan menjadi potongan-potongan kata, yang disebut sebagai token.
- *Stopwords Removal* adalah proses menghilangkan kata-kata yang kurang penting atau bisa menyesatkan dari kosakata, karena kata-kata tersebut tidak memiliki arti yang signifikan
- *Stemming* adalah mengurangi jumlah indikator yang berbeda dari informasi setelah ucapan yang diberi akhiran atau awalan ke bentuk dasarnya.

2.3 Labeling Data

Tahap berikutnya pelabelan pada kelas sentimen dengan menggunakan kategori positif, netral, dan negatif. Pada pelabelan ini menggunakan 2 tahapan yaitu google colab dan ahli bahasa. Data yang diperoleh dari hasil web scraping URL aplikasi Seabank di Google Play Store.

2.4 Splitting Data

Dalam penelitian ini menggunakan rasio pembagian 80 data training dan 20 data testing (80:20). Jumlah data latih menunjukkan berapa banyak sampel data yang digunakan untuk melatih model, sementara jumlah data uji menunjukkan berapa banyak sampel yang digunakan untuk menguji kinerja model [10]

2.5 Models

Pada tahap ini model di bagi menjadi 2 yaitu model *Support Vector Machine* (SVM) dan *Naïve Bayes*.

- *Support Vector Machine* (SVM)
Teknik Support Vector Machine (SVM) merupakan salah satu teknik pembelajaran learning dengan tingkat akurasi/presisi dan kualitas yang tinggi, menjadikannya algoritma yang sangat populer dibandingkan dengan algoritma lainnya. Namun harus melalui tahap pelatihan dan diuji sebelum diimplementasikan [11]. Support Vector Machine (SVM) adalah algoritma klasifikasi yang cepat dan sering digunakan karena menghasilkan hasil yang baik. Evaluasi performa menggunakan Confusion matrix. Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [12]
- *Naïve Bayes*
Naïve Bayes merupakan machine learning yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan Bayes. Penggunaan teorema Bayes pada algoritma Naïve Bayes adalah dengan menggabungkan prior probability dan conditional probability dalam suatu rumus yang dapat digunakan untuk menghitung probabilitas dari setiap kemungkinan klasifikasi [13].

2.6 Evaluasi

Evaluasi dalam penelitian ini bertujuan untuk mengetahui kinerja atau performa dari model yang diusulkan. Metode yang digunakan untuk evaluasi pada penelitian ini adalah confusion matrix [14]. Confusion matrix merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi prediksi dengan hasil klasifikasi yang sebenarnya. Tabel pada confusion matrix menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang salah diklasifikasikan. Berdasarkan nilai dari confusion matrix tersebut, dapat dihasilkan nilai akurasi, presisi, recall, dan f1-score dengan persamaan berikut [8].

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

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

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

$$F1 - Score = 2 \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

2.7 Word cloud

Pada tahap ini, memvisualisasikan data review berupa nilai dari kata-kata yang sering diberikan oleh pelanggan yang memberi ulasan pada aplikasi seabank [15]

3. HASIL DAN PEMBAHASAN

3.1. Hasil

3.1.1. Pengumpulan data ulasan

- Klasifikasi pada ulasan pengguna dengan SVM dan Naïve Bayes dapat menjadi pertimbangan bagi seabank sebagai perbaikan layanan dan mengkategorikan ulasan positif, negatif ataupun netral.

Berikut merupakan Script scrapping ulasan pada pemrograman dengan Python:

```
app_id = 'id.co.bankbkemobile.digitalbank'
start_date = datetime.date(2022, 1, 1)
end_date = datetime.date(2023, 12, 20)
```

Gambar 2. Scrapping URL id seabank di google play store

Banyaknya data yang diambil adalah 3804 data ulasan terbaru, Data ulasan diambil pada tanggal 01 januari 2022 – 20 Desember 2023. Selanjutnya data akan disimpan pada file yang berformat csv. Berikut contoh data ulasan seabank yang didapatkan dari Google Playstore:

Tabel 1. Hasil scrapping ulasan aplikasi seabank

No	At	Conten
1	01/12/2023 15:31:35	Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended 😞
2	17/12/2023 14:29:07	Sangat memudahkan untuk melakukan transaksi apapun, bukan hanya untuk transfer ke bank, tapi top up e-wallet juga ada , dan yang lebih mantap nya lagi ini gratis tanpa biaya admin . sukses terus seabank
3	19/11/2023 09:57:59	Harusnya sedia kan top up di indomaret biar lebih mudah.??

3.2 preprocessing data

Pada tahap *preprocessing* langkah awal yang dilakukan adalah tahap penghapusan duplikat, diikuti dengan beberapa langkah preprocessing data lainnya seperti *cleaning*, *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*. Berikut scrip program hapus data duplikat:

```
df.drop_duplicates(subset ="content", keep = 'first', inplace = True)

df.info()
```

Gambar 3. scrip program hapus data duplikat

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3789 entries, 0 to 3804
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   at           3789 non-null   object
1   userName     3789 non-null   object
2   score        3789 non-null   int64
3   content      3789 non-null   object
dtypes: int64(1), object(3)
memory usage: 148.0+ KB
```

Gambar 4. Membaca hasil hapus data duplikat

Hasil setelah menghapus data duplikat ulasan aplikasi sebank yang ada di google play store menggunakan bahasa pemrograman pyhton berjumlah 3789 dari 3804 ulasan. Kemudian data yang berjumlah 3789 akan masuk ke tahap *preprocessing data*.

1. *Cleaning*

Tabel 2. Hasil *cleaning*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended ☹️	Aplikasi gak ada GUNA Kalo terjadi kendala susah komplainnya Udah habis pulsa saya ribu lebih gak menemukan solusi CS ngomong lambat banget bertele tele masalah tidak terselesaikan Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya Tidak recommended

2. *Case folding*

Tabel 3. Hasil *case folding*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended ☹️	aplikasi gak ada guna kalo terjadi kendala susah komplainnya udah habis pulsa saya ribu lebih gak menemukan solusi cs ngomong lambat banget bertele tele masalah tidak terselesaikan buang buang waktu dan uang aja buat beli pulsa ngehubungin cs nya tidak recommended

3. *Tokenizing*

Tabel 4. Hasil *tokenizing*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recomended ☹️	[aplikasi,gak,ada,guna,kalo,terjadi,kendala,susah,komplainnya,udah,habis,pulsa,saya,ribu,lebih,gak,menemukan,solusi,cs,ngomong,lambat,banget,bertele,tele,masalah,tidak,terselesaikan ,buang,buang,waktu,dan,uang,aja,buat,beli,pulsa,ngehubungin,cs,nya,tidak,recommended]

4. *stopwords removal*

Tabel 5. Hasil *stopword removal*

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS	[aplikasi,gak,kalo,kendala,susah,komplainnya,udah,habis,pulsa,ribu,gak,menemukan,solusi,cs,ngomong,lambat,banget,bertele,tele,terselesa

ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recommended 😞	ikan,buang,buang,uang,aja,beli,pulsa,ngehubu ngin,cs,nya,recommended]
---	---

5. stemming

Tabel 6. Hasil stemming

Sebelum	Sesudah
Aplikasi gak ada GUNA. Kalo terjadi kendala susah komplainnya. Udah habis pulsa saya 30ribu lebih gak menemukan solusi. CS ngomong lambat banget, bertele tele, masalah tidak terselesaikan !!! Buang buang waktu dan uang aja buat beli pulsa ngehubungin CS nya !!!! Tidak recommended 😞	aplikasi gak kalo kendala susah komplain udah habis pulsa ribu gak temu solusi cs ngomong lambat banget tele tele selesai buang buang uang aja beli pulsa ngehubungin cs nya recommended

3.3 Labeling Data

Labeling data menggunakan google colab dan ahli bahasa. 1895 Data yang akan diberikan label pada google colab dan 1894 data yang diggunakan pada ahli bahasa, dari Total keseluruhan data sebanyak 3.789 ulasan. sentimen dikategorikan menjadi positif, netral dan negatif.

1. Labeling google colab

Hasil pelabelan sentimen google colab menunjukkan bahwa 438 data memiliki sentimen positif, 1379 data memiliki sentimen netral, dan 77 data memiliki sentimen negatif. Berdasarkan hasil pelabelan menggunakan ahli bahasa, dapat disimpulkan bahwa sentimen pengguna aplikasi Seabank pada rentang waktu 01 Januari 2022 hingga 20 Desember 2023 didominasi oleh sentimen netral dari keseluruhan data sebanyak 1894 ulasan.

bantu mudah cepat puasaaassssss	netral
seabank mudah belanja pokok is the best	positif
transaksi cepat ribet	netral
cerita nyah no gua udah gak aktif gagal verifikasi lapor kek gtu ajah susah banget ber belit belit ngirim email dah males	negatif
gimana ya transfer suruh verifikasi wajah ulang kali gabisa	netral
mudah smoga depan	netral
mudah lancar bantu	netral
bagus tawar tarik	netral
mantap gampang ga ribed	netral
puas pakai seabank apa mudah	netral
gratis top up ewallet nya manfaat	positif
senang seabank simpel ngga ribet	netral
bagus enak aplikasi ny simple garibet	netral
mudah pakai gratis biaya admin	positif
membantugratis tranfer pulasukses seabank	netral
aplikasi seabankn bantu nyesal aplikasi bagus banget	netral
lumayan tf admin bunga mantap	netral
aplikasi rusak telpon cs	netral
bantu isi ewallet transfer bank gratis x bulan	positif
bantu gratis biaya admin	positif

Gambar 5. Hasil labeling google colab

2. Labeling ahli bahasa

Hasil pelabelan sentimen ahli bahasa menunjukkan bahwa 1100 data memiliki sentimen positif, 308 data memiliki sentimen netral, dan 486 data memiliki sentimen negatif. Berdasarkan hasil pelabelan menggunakan ahli bahasa, dapat disimpulkan bahwa sentimen pengguna aplikasi Seabank pada rentang waktu 01 Januari 2022 hingga 20 Desember 2023 didominasi oleh sentimen positif.

sy isi saldo seabank checkout sophee klaim voucher seabank bayar gagal ulang sy coba kali kali tdk saldondi seabank sy r negatif aplikasi force closeaplikasi update baru hp direset bla bla bla force closeit ny g benerin mau invest jg aplikasi g negatif ga aplikasi shopee kendala suruh hubung cs seabank aplikasi aplikasi dar login gabisa reset password masuk password t negatif suka aplikasi top up transfer nya cepathanya tunggu detik saldo masuksangat nyaman dompet suka transfer biaya admn positif kirim gratis bank mantap pokoke mohon tambahn fitur tarik tunai alfamart indomart guna pindah aplikasi bank belah t positif aplikasi kemarin bagus apl login gak tf bayar sekolah hadeueueueuh mahon bantu dooong baik kasih bintang nati kcalc netral sambung shopee usaha ragu ya lihat reviewnya g bagus kendala komplain nya buka deh netral mohon maaf naruh bintang notice baca sk deposito nya kerl salah satu sewaktuwaktu bijak ubah modifikasi nonaktif he negatif mantap serba cepat efisien nyama banget dipake saran sih kl buka akun rekening kena biaya tambah positif bunga simpan bunga deposito bunga simpan pa bunga deposito maksimal pa turun promo transfaran kecewa udah bair negatif gin wir kasih mnt kode otp sms kode otp nya lot banget menit keburu kadaluarsa gak log in jamkrna minta otpsistem bu negatif aman betah aplikasi seabank transaksi yg mudah gratis bank top up ewallet jg gratis bunga yg hari saran tolong adu banr positif beli barang dishoppie menggunakan bayar seabank dpt chasback beli shoppie gak chasbacknya yaaharusnya tp byr ny netral vertifikasi wajah tolong sensitifmasa geser dikit banget s semut gagal aplikasi geser dikit gak ngaruh mah tahan nafas b netral salah password klo blokir transaksi dgn nominal buka rek pakai hp klo lupa salah rek ga bs buka lg dgn alas tdk sesuai ktj negatif share alam guna barusejauh gabung alhamdulillah aman kondusif hasil ajak teman selesai misi bonus saldo k beneran rr positif aplikasi baik aplikasi bank indonesia seabank baik nggak ribet beliit beliit guna gampang nggk ngeleg luuuuar rekomend positif transfer mudah tolong adakan search name yg transfer mudah ga scroll terimakasih smoga kembang positif

Gambar 6. Labeling ahli bahasa

3.4 Spitting Data

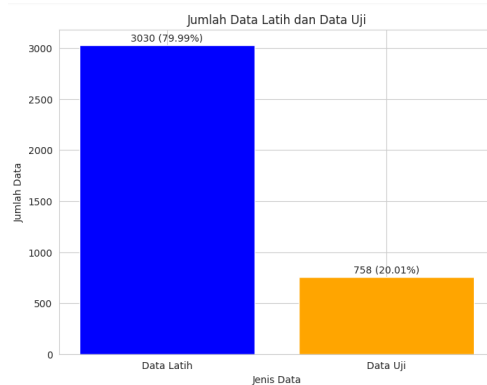
Dalam penelitian ini menggunakan rasio pembagian 80 data training dan 20 data testing (80:20).

```
from sklearn.model_selection import train_test_split

# Bagi data menjadi data pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(data['steming_data'], data['sentiment'], test_size=0.2, random_state=42)
```

Gambar 7. Program data latih dan data uji

Hasil pembagian data training dan data testing dengan rasio (80:20) menunjukkan bahwa sebanyak 3031 ulasan (atau 79.99% dari total data) ditetapkan sebagai data latih, sementara 758 ulasan (atau 20.01% dari total data) ditetapkan sebagai data uji dari keseluruhan 3788 ulasan. Dengan pembagian yang proporsional antara data latih dan data uji, diharapkan dapat memastikan evaluasi yang akurat terhadap kinerja model dalam menganalisis sentimen ulasan pengguna aplikasi Seabank. Berikut hasil pembagian data yang menampilkan jumlah data dengan rasio 80 data training dan 20 data testing (80:20)



Gambar 8. Hasil pembagian data latih dan data uji

3.5 Models

1. Naïve Bayes

Dalam penelitian ini, digunakan tiga jenis model Naïve Bayes, yaitu model Gaussian Naïve Bayes, Multinomial Naïve Bayes, dan Bernoulli Naïve Bayes.

```
import numpy as np
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.naive_bayes import GaussianNB, MultinomialNB, BernoulliNB
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import LabelEncoder
from sklearn.pipeline import make_pipeline
from sklearn import metrics
```

Gambar 9. Model Naïve bayes

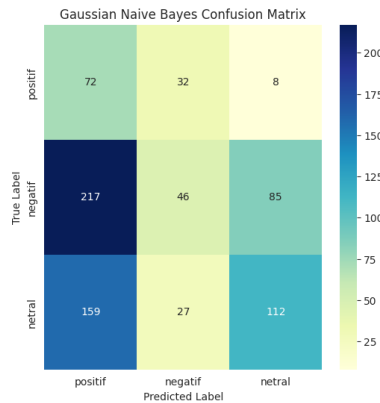
Hasil perbandingan menunjukkan bahwa perolehan nilai akurasi tertinggi ditemukan pada model Bernoulli Naïve Bayes dengan akurasi mencapai 61.1%. Sementara itu, Multinomial Naïve Bayes mencapai akurasi sebesar 58.3%, sedangkan model Gaussian Naïve Bayes memperoleh akurasi terendah dengan nilai 30.3%.

Dari hasil akurasi, dapat disimpulkan bahwa dalam konteks analisis sentimen ulasan pengguna aplikasi Seabank, model Bernoulli Naïve Bayes memberikan kinerja yang lebih baik dalam memprediksi sentimen pengguna dibandingkan dengan model Multinomial Naïve Bayes dan Gaussian Naïve Bayes.

Tabel 7. Hasil Akurasi naïve bayes

No	Model	Akurasi
1	Gaussian naïve bayes	30.3%
2	Multinomial naïve bayes	58.3%
3	Bernoulli naïve bayes	61.1%

a. Gaussian naïve bayes



Gambar 10. Confusion matrix gaussian naïve bayes

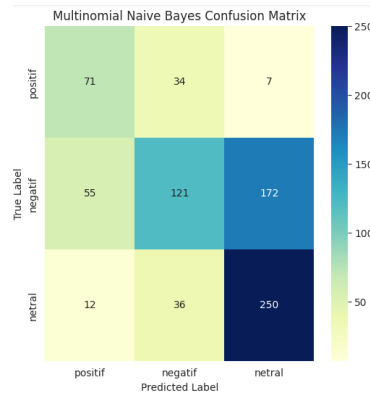
Berhasil mengklasifikasikan secara benar pada data positif (True Positive) sebesar 72, data benar negatif (True Negative) sebanyak 46 dan data benar netral (True Neutral) sebanyak 112. Sedangkan False negatives (FN) 32 False positives (FP) 8 . Setelah pengujian menggunakan confusion matrix, langkah selanjutnya adalah menghitung nilai akurasi, presisi, recall, dan F1-score.

Tabel 8. Hasil evaluasi gaussian naïve bayes

	Precision	Recall	F1-score	Support
Negatif	0.16	0.64	0.26	122
Netral	0.20	0.13	0.20	348
Positif	0.55	0.38	0.45	298
Akurasi			0.30	758

Dari tabel diatas, hasil Presisi untuk kelas negatif adalah 0.16, untuk kelas netral adalah 0.44, dan untuk kelas positif adalah 0.55. Recall untuk kelas negatif adalah 0.64, untuk kelas netral adalah 0.13, dan untuk kelas positif adalah 0.38. F1-score untuk kelas negatif adalah 0.26, untuk kelas netral adalah 0.20, dan untuk kelas positif adalah 0.45. dan Akurasi sebesar 0.30 menunjukkan bahwa model secara keseluruhan benar dalam 30% prediksinya.

b. Multinomial naïve bayes



Gambar 11. Confusion matrix multional naïve bayes

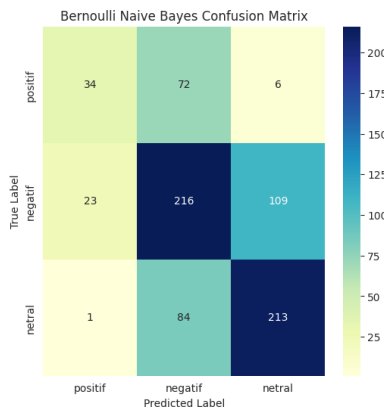
Berhasil mengklasifikasikan secara benar pada data positif (True Positive) sebesar 71, data benar negatif (True Negative) sebanyak 121 dan data benar netral (True Neutral) sebanyak 250. Sedangkan False negatives (FN) 34 False positives (FP) 7. Setelah pengujian menggunakan confusion matrix, langkah selanjutnya adalah menghitung nilai akurasi, presisi, recall, dan F1-score.

Tabel 9. Hasil evaluasi multional naïve bayes

	Percission	Recall	F1-score	Support
Negatif	0.51	0.63	0.57	122
Netral	0.63	0.35	0.45	348
Positif	0.58	0.84	0.69	298
Akurasi			0.58	758

Dari tabel diatas, hasil Presisi untuk kelas negatif adalah 0.51, untuk kelas netral adalah 0.63, dan untuk kelas positif adalah 0.58. Recall untuk kelas negatif adalah 0.63, untuk kelas netral adalah 0.35, dan untuk kelas positif adalah 0.84. F1-score untuk kelas negatif adalah 0.57, untuk kelas netral adalah 0.45, dan untuk kelas positif adalah 0.69. Akurasi sebesar 0.58 menunjukkan bahwa model secara keseluruhan benar dalam 58% prediksinya.

c. Bernoulli naïve bayes



Gambar 12. Confusion matrix bernoulli naïve bayes

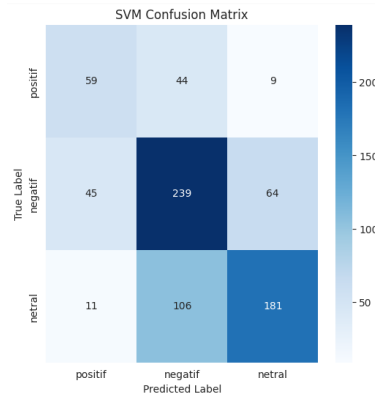
Berhasil mengklasifikasikan secara benar pada data positif (True Positive) sebesar 34, data benar negatif (True Negative) sebanyak 216 dan data benar netral (True Neutral) sebanyak 213. Sedangkan False negatives (FN) 72 False positives (FP) 6. Setelah pengujian menggunakan confusion matrix, langkah selanjutnya adalah menghitung nilai akurasi, presisi, recall, dan F1-score.

Tabel 10. hasil evaluasi bernoulli naïve bayes

	Percission	Recall	F1-score	Support
Negatif	0.59	0.65	0.40	122
Netral	0.58	0.62	0.60	348
Positif	0.65	0.71	0.68	298
Akurasi			0.61	758

Dari tabel diatas, Presisi untuk kelas negatif adalah 0.59, untuk kelas netral adalah 0.58, dan untuk kelas positif adalah 0.65. Recall untuk kelas negatif adalah 0.65 untuk kelas netral adalah 0.62, dan untuk kelas positif adalah 0.71. F1-score untuk kelas negatif adalah 0.40, untuk kelas netral adalah 0.60, dan untuk kelas positif adalah 0.68. Akurasi sebesar 0.61 menunjukkan bahwa model secara keseluruhan benar dalam 61% prediksinya.

2. Support Vectort Machines (SVM)



Gambar 13. Confusion matrix SVM

Berhasil mengklasifikasikan secara benar pada data positif (True Positive) sebesar 59, data benar negatif (True Negative) sebanyak 219 dan data benar netral (True Neutral) sebanyak 181. Sedangkan False negatives (FN) 41 False positives (FP) 9. Setelah pengujian menggunakan confusion matrix, langkah selanjutnya adalah menghitung nilai akurasi, presisi, recall, dan F1-score.

Tabel 11. Hasil evaluasi SVM

	Percission	Recall	F1-score	Support
Negatif	0.51	0.53	0.52	122
Netral	0.61	0.69	0.65	348
Positif	0.671	0.61	0.66	298
Akurasi			0.63	758

Dari tabel diatas, Presisi untuk kelas negatif adalah 0.51, untuk kelas netral adalah 0.61, dan untuk kelas positif adalah 0.71. Recall untuk kelas negatif adalah 0.53, untuk kelas netral adalah 0.69, dan untuk kelas positif adalah 0.61. F1-score untuk kelas negatif adalah 0.52, untuk kelas netral adalah 0.65, dan untuk kelas positif adalah 0.66. Akurasi sebesar 631 menunjukkan bahwa model secara keseluruhan benar dalam 63% prediksinya.

3.6 Perbandingan model naïve bayes dan SVM

```
import matplotlib.pyplot as plt

# Data akurasi Naive Bayes
naive_bayes_labels = ['GNB', 'MNB', 'BNB']
naive_bayes_accuracies = [metrics.accuracy_score(y_test, y_pred_gnb),
                           metrics.accuracy_score(y_test, y_pred_mnb),
                           metrics.accuracy_score(y_test, y_pred_bnb)]
```

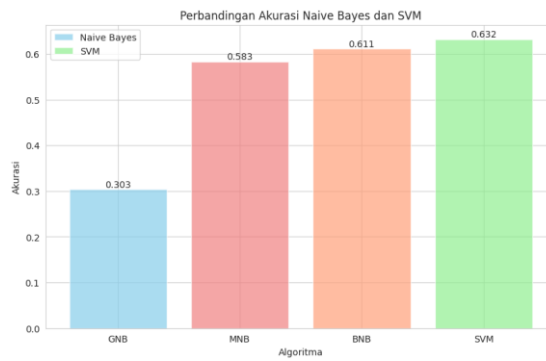
Gambar 14. Program Naïve bayes

```
# Data akurasi SVM
svm_label = 'SVM'
svm_accuracy = accuracy_score(y_test, y_pred_svm)

# Plotting
plt.figure(figsize=(10, 6))
```

Gambar 15. Program SVM

Berikut merupakan hasil perbandingan akurasi tiga model naïve bayes dan model SVM:



Gambar 16. Hasil perbandingan model svm dan naïve bayes

Hasil dari perbandingan kinerja antara dua jenis algoritma klasifikasi, yaitu Naïve Bayes (Multinomial, Bernoulli, dan Gaussian) dan Support Vector Machines (SVM). Kinerja SVM menunjukkan bahwa nilai akurasi tertinggi ditemukan, dengan nilai akurasi mencapai 63.2%. Sementara itu, nilai akurasi tertinggi dari Naïve Bayes terdapat pada model Bernoulli Naïve Bayes, dengan nilai akurasi 61.1%. Namun, perlu dicatat bahwa nilai akurasi terendah ditemukan pada model Gaussian Naïve Bayes, dengan nilai akurasi sebesar 0.303, sedangkan akurasi Multinomial Naïve Bayes adalah 58.3%. Dari hasil analisis ini, dapat disimpulkan bahwa dalam konteks analisis sentimen ulasan pengguna aplikasi Seabank, model Support Vector Machines (SVM) memberikan kinerja yang lebih baik dalam memprediksi sentimen pengguna dibandingkan dengan model Naïve Bayes. Namun, dari tiga tipe Naïve Bayes yang diuji, Bernoulli Naïve Bayes menunjukkan kinerja tertinggi, sedangkan Gaussian Naïve Bayes menunjukkan kinerja terendah.

3.7 WordCloud

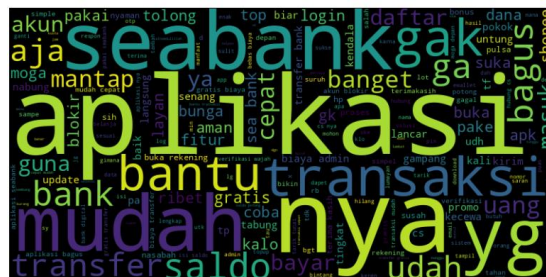
```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Pisahkan dataset berdasarkan sentimen
sentimen_negatif = data[data['sentiment'] == 'negatif']['steming_data'].str.cat(sep=' ')
sentimen_positif = data[data['sentiment'] == 'positif']['steming_data'].str.cat(sep=' ')
sentimen_netral = data[data['sentiment'] == 'netral']['steming_data'].str.cat(sep=' ')

```

Gambar 17. Program wordcloud

Berikut ini merupakan gambar wordcloud sentimen gabungan negatif, netral dan positif:



Gambar 18. WordCloud aplikasi seabank

Berdasarkan visualisasi, dapat disimpulkan bahwa kata-kata "aplikasi", "mudah", dan "Seabank" merupakan tiga kata yang paling sering dibicarakan oleh pengguna aplikasi Seabank. Hal ini menunjukkan bahwa fokus utama dalam ulasan pengguna cenderung berpusat pada aplikasi Seabank itu sendiri dan kemudahan penggunaannya. Selain itu, terdapat pula kata-kata lain yang muncul dengan frekuensi yang signifikan, seperti "nya", "transaksi", "yg" (singkatan dari "yang"), dan "saldo". Kemunculan kata-kata ini menunjukkan aspek-aspek tertentu dari aplikasi yang sering menjadi perbincangan oleh pengguna, seperti fitur-fitur transaksi, informasi saldo, dan penggunaan aplikasi secara umum.

Informasi ini memberikan wawasan yang berharga bagi aplikasi Seabank dalam memahami pandangan, kebutuhan, dan pengalaman pengguna terhadap aplikasi mereka. Dengan memperhatikan kata-kata yang sering muncul dalam ulasan pengguna, Seabank dapat mengidentifikasi kekuatan dan kelemahan aplikasi, serta membuat perbaikan atau peningkatan yang sesuai.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka kesimpulan yang dapat diambil adalah:

1. Algoritma klasifikasi Support Vector Machine dan naïve bayes dapat diterapkan dalam kontesks ulasan pengguna aplikasi seabank.
2. Berdasarkan klasifikasi sentimen google colab dan ahli bahasa, didapatkan bahwa terdapat tiga kelas sentimen, yaitu positif, netral, dan negatif. ini menunjukkan bahwa mayoritas ulasan pengguna aplikasi Seabank pada google colab cenderung netral dan ahli bahasa cenderung positif.
3. Hasil perbandingan akurasi model SVM dan naïve bayes
Dari hasil analisis, terlihat bahwa model SVM mencapai akurasi tertinggi dengan persentase 63.2%, Sementara itu, model Gaussian Naïve Bayes menunjukkan akurasi yang jauh lebih rendah, hanya mencapai 30.3%. Dengan demikian, dapat disimpulkan bahwa model SVM lebih efektif dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Seabank daripada model Naïve Bayes.

UCAPAN TERIMAKASIH

Penulis ingin mengucapkan terimakasih yang tak terhingga kepada orangtua, dosen pembimbing, serta semua yang telah memberikan bantuan dan dukungan dalam menyelesaikan penelitian dan publikasi ini.

REFERENCES

- [1] S. N. N. G. T. I. Alun Sujjadaa, ""analisis sentimen terhadap review bank digital menggunakan metode support vector machine (SVM)"" , *rekayasa teknologi nusa putra*, vol. Vol 9 No 2 (2023), 2023.
- [2] B. M. C. H. F. J. J. L. A. N. F. Jesslyn, ""Analisis Strategi Bisnis dan Tingkat Persaingan dengan Kompetitor pada Perusahaan Seabank dalam Dunia Perbankan"" , *Jurnal Mirai Management*, Vols. Vol 8, No 2 (2023), 2023.
- [3] I. G. A. p. a. L. K. A. S. S. T. W. N. S. P. A. Erni Febrina Harahap, ""Digital bank transformation: A content analysis of SEABank"" , *komunikasi profesional*, vol. Vol. 7 No. 3 (2023), 2023.
- [4] A. F. G. D. Asep Saepudin, ""Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee"" , *teknokompak*, Vols. Vol 18, No 1 (2024), 2024.
- [5] e. w. Noviah Dwi Putranti, ""Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector machine"" , *IJCCS (indonesia jurnal of computing and cybernetics system)*, Vols. Vol 8, No 1.
- [6] T. S. S. Y. A. Mujaddid Izzul Fikri, ""perbandingan metode naive bayes dan support vector machine pada analisis sentimen twitter"" , *STIKI Informatika Jurnal*, vol. Vol 10 No 02 (2020), 2020.
- [7] L. B. I. d. M. A. Mudeb, ""perbandingan metode klasifikasi support vector machine dan naive bayes untuk analisis sentimen ulasan tekstual di google play store"" , *ILKOM Jurnal Ilmiah* , vol. Vol. 12 No. 2, 2020.
- [8] A. F. S. F. Friska Aditia Indriyani, ""Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine"" , *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. Vol 10 No 2 (2023), 2023.
- [9] a. U. V. f. e. R. W. samsir, "analisis sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *media informatika budidarma*, Vols. vol 5, Nomor 1, pp. 157-163, 2021.
- [10] G. A. L. B. A. W. D. S. P. Bobby Kurniadi W. Hariyanto Prasetyo, ""Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah"" , *SENAMIKA*, Vols. Vol 2, No 2 (2021), 2021.

- [11] F. S. Selva Indah Nurhafida, "Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM)," *J-SAKTI (sains komputer informatika)*, vol. VOL 6 NO 1 9 (2022), Maret 2022.
- [12] S. A. P. dwi normawati, ""Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen berbasis teks pada twiiter"," *J-SAKTI (sains komputer dan informatika)*, Vols. vol 5, no 2 (2021), 2021.
- [13] E. D. N. M. H. A. winda yulita, ""Analisis Sentimen Terhadap Opini Masyarakat Tentang vaksin covid-19 menggunakan algoritma naive bayes Classifier"," *E-Journal Universitas Teknokrat Indonesia*, Vols. Vol. 2, No. 2, 2021, pp. 1-9, 2021.
- [14] G. F. N. W. E. S. Dianati Duei Putri, ""analisis sentimen kinerja dewan perwakilan rakyat (DPR) pada twitter menggunakan metode niave bayes classifier"," *jurnal informatika dan teknik elektro tarapan*, Vols. Vol 10, No 1 (2022), 2022.
- [15] J. J. A. Limbong, ""Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naïve Bayes dan K-Nearest Neighbor"," *satya wacana institutional repository*, 2022.