

Analisis Sentimen BRImo dan BCA Mobile Menggunakan Support Vector Machine dan Lexicon Based

Muhamad Rizky Pratama^{1*}, Yudhi Raymond Ramadhan², Mutiara Andayani Komara³
 Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Purwakarta, Indonesia
 *e-mail *Corresponding Author*: muhamadrizky61@wastukencana.ac.id

Abstract

Services available on mobile banking can make transactions and request financial information, such as checking balances, viewing account mutation history and the like even so, it is felt that there are still opinions and complaint submitted by users in the Play Store review column of the application. Based on these problems, sentiment analysis research was carried out on reviews of the BRImo and BCA Mobile applications on the Play Store as research objects using the Support Vector Machine and Lexicon Based classification algorithms. Based on this result, it is known that the BCA Mobile application has more negative sentiment than the BRImo application. The sentiment classification of application reviews in testing the Support Vector Machine and Lexicon Based on the BRImo application obtained 94% accuracy and, on the BCA, Mobile application the accuracy was 95%

Keywords: *Lexicon Based; Mobile Banking; Play Store; Sentiment Analysis; Support Vector Machine*

Abstrak

Layanan yang ada pada *mobile banking* dapat melakukan transaksi dan meminta informasi keuangan, seperti memeriksa saldo, melihat riwayat mutasi rekening, dan sejenisnya. Meskipun demikian, dirasa masih terdapat pendapat dan keluhan yang disampaikan oleh pengguna pada kolom ulasan *play store* dari aplikasi tersebut. Berdasarkan permasalahan tersebut dilakukan penelitian analisis sentimen pada ulasan aplikasi BRImo dan BCA Mobile di *Play Store* sebagai objek penelitian menggunakan algoritma klasifikasi *Support Vector Machine* dan *Lexicon Based*. Berdasarkan hasil tersebut diketahui bahwa aplikasi BCA Mobile memiliki sentimen negatif yang lebih banyak dibandingkan aplikasi BRImo. Klasifikasi sentimen ulasan aplikasi dalam pengujian menggunakan *Support Vector Machine* pada aplikasi BRImo didapatkan *accuracy* 94%, Lalu pada aplikasi BCA Mobile *accuracy* 95%.

Kata kunci: *Analisis Sentimen; Lexicon Based; Mobile Banking; Play Store; Support Vector machine*

1. Pendahuluan

Meningkatnya kemajuan teknologi informasi dan komunikasi di berbagai sektor seperti pendidikan, kesehatan, perbankan dan berbagai sektor lainnya. Pada era digitalisasi ini terutama pada sektor perbankan sudah banyak perusahaan bank yang memiliki aplikasi *mobile banking*. Layanan yang ada pada *mobile banking* dapat melakukan transaksi dan meminta informasi keuangan, seperti memeriksa saldo, melihat riwayat mutasi rekening dan sejenisnya tanpa perlu di mesin ATM ataupun di bank. Berbagai macam upaya dilakukan oleh perbankan dalam mengalami dampak kemajuan *e-bussiness* dengan banyak berinovasi dalam memanfaatkan perkembangan teknologi informasi dan komunikasi. Dalam memberikan servis terbaik dan memaksimalkan kepuasan pengguna/nasabah setia perbankan mulai mengembangkan solusi yang tepat sesuai kebutuhan pengguna dan loyal mengaplikasikan inovasi teknologi digital *banking* di tengah ramainya transformasi digital.

Namun nyatanya dibalik kemudahan tersebut dirasa masih banyak pendapat atau keluhan yang disampaikan oleh pengguna yaitu suatu bentuk ketidaknyamanan kondisi tersebut dapat mengurangi kepercayaan konsumen pada layanan tersebut. Adanya persaingan antar bank, salah satu bank pemerintah (BUMN) terbesar yaitu PT. Bank Rakyat Indonesia (persero) Tbk, dengan bank swasta terbesar yakni PT. Bank *Central Asia* Tbk, terkait kualitas

layanan, merupakan salah satu tantangan utama yang dialami dalam menjaga kesetiaan pengguna/nasabah. Dilihat pada *review* atau ulasan BRImo dan BCA *Mobile* di *Play Store*, mengutarakan bahwa pengguna/nasabah sulit untuk melakukan verifikasi wajah (aktivasi *login*), transaksi, juga ditemukan kegagalan *transfer* tetapi saldo berkurang. Hal tersebut membuktikan dari banyaknya tanggapan buruk yang disampaikan oleh pengguna/nasabah. Kejadian tersebut menunjukkan bahwa adanya ketidaksesuaian antara keadaan sebenarnya dengan ekspektasi layanan berkualitas yang diharapkan bagi pengguna.

Analisis sentimen adalah sebuah metode untuk mendapatkan opini atau pandangan pengguna tentang beberapa tema atau teks [1]. Tugas dasar analisis ini yaitu membagi sentimen positif dan negatif berdasarkan ulasan yang disampaikan oleh pengguna aplikasi *mobile banking* (BRImo dan BCA *Mobile*) di *Play Store*. Kemudian pada proses *labelling* menggunakan *Lexicon Based*. *Lexicon Based* adalah kamus yang berisi kata-kata Bahasa Indonesia dan pada tiap kata memiliki bobot untuk menentukan polaritas dalam menentukan jenis sentimen positif atau negatif pada opini [2]. Kemudian dilakukan *preprocessing data*, yaitu proses mengubah raw data menjadi data terstruktur dengan diolah oleh bahasa pemrograman [3]. Selanjutnya dilakukan tahap *split data* dan pembobotan kata menggunakan TF-IDF. TF-IDF yaitu proses mengubah teks menjadi data numerik pada setiap kata yang ada pada dokumen [4]. Proses klasifikasi pada penelitian ini menggunakan algoritma *Support Vector Machine*. *Support Vector Machine* adalah salah satu metode klasifikasi dengan menggunakan metode *supervised learning* yang memprediksi kelas berdasarkan hasil proses data latih SVM mampu memisahkan data dengan baik dibandingkan algoritma klasifikasi lainnya [5]. Lalu dievaluasi menggunakan *confusion matrix*. *confusion matrix* yaitu proses evaluasi model klasifikasi untuk mengukur *accuracy*, *precision* dan *recall* dari prediksi dari model yang telah dibuat [6]. Proses evaluasi menggunakan *Microsoft Power BI* memvisualisasikan *stacked column chart*, *pie chart*, *clustered column chart*, dan *WordCloud*

Penelitian ini dilakukan bertujuan untuk menganalisis sentimen ulasan aplikasi *mobile banking* (BRImo dan BCA *Mobile*) di *Play Store* dengan mengelompokkan sentimen positif dan negatif menggunakan *Lexicon Based* dan Algoritma *Support Vector Machine* agar pihak bank dapat mengetahui sentimen pengguna terhadap aplikasi *mobile banking* tersebut. Serta dapat membantu mengevaluasi juga sebagai acuan dalam meningkatkan layanan aplikasi *mobile banking* kedepannya.

2. Tinjauan Pustaka

Sebuah penelitian yang telah dilakukan sebelumnya oleh Doloksaribu & Samuel (2022), yang berjudul “Komparasi Algoritma *Data Mining* untuk Analisis Sentimen Aplikasi PeduliLindungi”. Penelitian ini dilakukan untuk mengkomparasi dua algoritma yakni *Naïve Bayes* dan *Support Vector Machine*. Hasilnya menunjukkan bahwa algoritma *Support Vector Machine* menjadi algoritma yang terbaik dalam tingkat akurasi sebesar 89,05% dengan jumlah data 1.948 dan perbandingan data yang bukan 1:1 pada masing-masing label bukan halangan bagi algoritma SVM dalam meningkatkan performanya dengan data yang lebih banyak [7].

Penelitian selanjutnya “Analisis Sentimen *Zoom Cloud Meetings* di *Play Store* menggunakan *Naïve Bayes* dan *Support Vector Machine*” yang dilakukan oleh Herlinawati et al (2020). Hasil penelitian ini dengan membandingkan dua algoritma tersebut dari hasil evaluasi bahwa *Support Vector Machine* lebih unggul ketimbang *Naïve Bayes* dengan 1.007 jumlah dataset dengan akurasi sebesar [8].

Penelitian sebelumnya dengan judul “Analisis Sentimen Aplikasi BRImo pada Ulasan Pengguna di *Google Play* Menggunakan Algoritma *Naïve Bayes*” hasil penelitian terhadap analisis sentimen pengguna aplikasi BRImo dengan hasil scraping data ulasan dari bulan November 2022 – Januari 2023 sebanyak 1550 data dengan proses *labelling* dilakukan secara manual didapatkan hasil sentimen positif sebanyak 1012 atau 53,09% dan sentimen negatif 894 atau 46,90% dan hasil klasifikasi menggunakan *Naïve Bayes* didapatkan *accuracy* 84,52% [9].

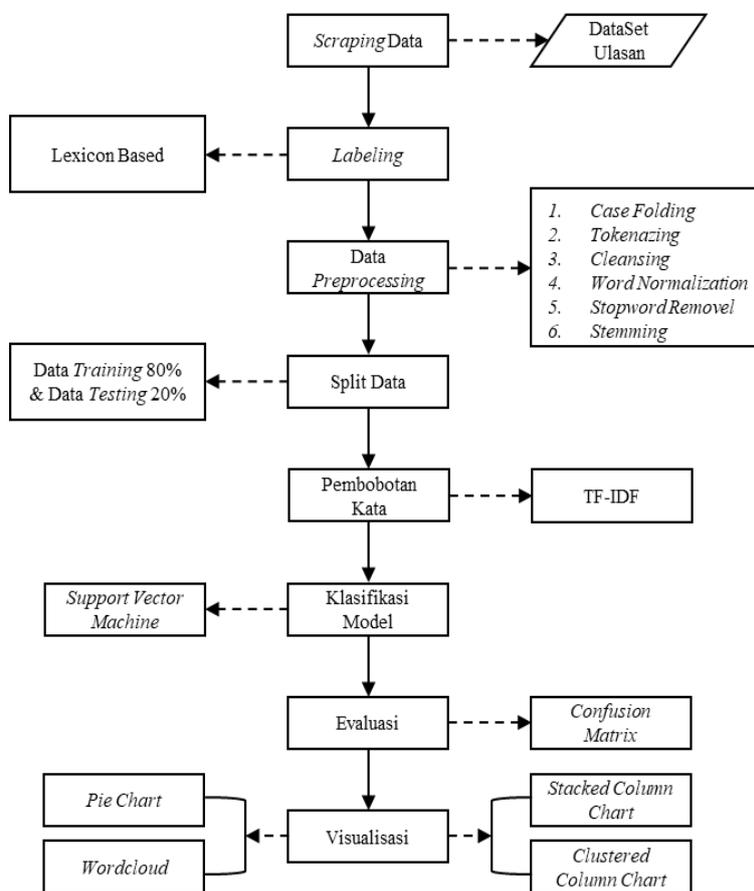
Pada penelitian sebelumnya dengan judul “Analisis Sentimen *Review* Pengguna BCA *Mobile* Menggunakan Teks *Mining*” Jumlah data ulasan yang didapatkan sebanyak 500 data yang didapatkan menggunakan metode web *scrapping* pada Januari 2022 – September 2022 dengan pemberian label berdasarkan *score/rating*. *Rating* 1-2 diberi label negatif, sedangkan 4-6 diberi label positif. Maka dari itu didapatkan data label positif maupun negatif sebanyak 250 data pada masing-masing label, serta hasil dari klasifikasi menggunakan *Naïve Bayes* mendapatkan nilai akurasi sebesar 82% [10].

Selain itu, terdapat penelitian berjudul “Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi menggunakan Pendekatan *Lexicon Based Features* dan *Support Vector Machine*. Metode SVM dan *Lexicon Based* memperoleh *accuracy* 0.6, dibandingkan tanpa *Lexicon Based* yang mendapatkan *accuracy* 0.48. Dapat disimpulkan bahwa *Lexicon Based* dapat mengoptimalkan nilai evaluasi sebesar 12% [11].

Berdasarkan penelitian yang telah dilakukan sebelumnya, khususnya pada analisis ulasan aplikasi, belum ada yang memfokuskan analisis sentimen pada perbandingan dua aplikasi *mobile banking*, yaitu BRI_{mo} dan BCA *Mobile*, menggunakan algoritma *Support Vector Machine*. Oleh karena itu, dalam penelitian ini, peneliti melakukan pengelompokan sentimen terhadap kedua aplikasi tersebut menggunakan *Lexicon Based* dan melakukan model klasifikasi *Support Vector Machine* dan *confusion matrix* untuk menilai performa model dengan *accuracy*, *precision*, dan *recall*.

3. Metodologi

Dalam penelitian ini, terdapat beberapa langkah yang dijalankan, mulai dari pengambilan data (*scraping*), *labelling*, *data preprocessing*, pembobotan kata (TF-IDF), data *training* dan data *testing*, klasifikasi *support vector machine*, evaluasi (*confusion matrix*), dan visualisasi. Rancangan penelitian yang akan dilakukan sebagaimana ditunjukkan pada gambar 2 berikut ini:



Gambar 1 Kerangka Penelitian

1) Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil data ulasan pada Aplikasi *mobile banking* BRI_{mo} dan BCA *Mobile* di *Play Store* berbahasa Indonesia. Data diambil menggunakan teknik *scraping* dengan bantuan *library google-play-scrapers* menggunakan *Python* [12]. Data ulasan diambil pada 10 April 2023 dengan rentang waktu ulasan sejak (5 Oktober 2022 – 9 April 2023) dengan penentuan kategori *most_relevant* jumlah data yang diambil berjumlah 2.000 data pada masing-masing aplikasi atau total 4.000 data ulasan secara keseluruhan.

2) *Labelling Lexicon Based*

Langkah selanjutnya yaitu melabeli data untuk menentukan sentimen atau polaritas dari ulasan yang diperoleh. Pendekatan *Lexicon Based* merupakan kamus yang berisi kata berbahasa Indonesia yang biasanya digunakan untuk proses pelabelan. Nilai polaritas berdasarkan score, bila $score > 0$ yaitu positif, bila $score = 0$ yaitu netral, dan bila $score < 0$ yaitu negatif [2].

3) *Preprocessing Data*a) *Case Folding*

Operasi ini akan melakukan perubahan seluruh huruf/teks menjadi huruf *lowercase*/huruf kecil [13], karena data yang didapat menggunakan huruf yang beragam seperti huruf besar dan kecil.

b) *Cleansing*

Pada tahap ini melakukan eliminasi karakter *non alfabetis* untuk menurunkan *noise* seperti tanda baca titi (.), koma (,) dan tanda tanya (?) serta simbol dan emotikon [14].

c) *Tokenizing*

Kalimat yang ada pada ulasan akan dilakukan pemotongan kalimat menjadi potongan kata demi kata dari kalimat panjang [15]. Tujuan dari proses *tokenizing* yakni untuk membuat nilai bobot setiap kata menjadi lebih mudah.

d) *Word Normalization*

Digunakan untuk memperbaiki kalimat *slang* atau tidak baku menjadi kalimat yang baik dan benar sesuai dengan kaidah tata bahasa Indonesia. Tahap ini dilakukan dengan bantuan kamus *slangword* dengan memeriksa kata yang ada pada ulasan termuat dalam kamus *slangword* atau tidak. Apabila kata tidak resmi tercantum dalam kamus, maka kata tersebut akan diganti menjadi kata resmi yang termuat pada kamus [16].

e) *Stopword Removal*

Cara ini dilakukan untuk menghilangkan kata yang tidak berhubungan, kata yang tidak mempunyai makna seperti kata konjungsi/penghubung “ke”, “agar” atau “yang” dan kata-kata yang tidak terkait dengan kata sifat [3].

f) *Stemming*

Proses pemetaan mengubah sebuah kata menjadi kata dasar dengan menghapus semua afiks yang terdiri dari awalan dan akhiran sehingga dapat memperkecil panjang string [17].

4) *Split data*

Proses *split* data yaitu pemisahan *dataset* ulasan membuat data *training* (pelatihan) dan data *testing* (pengujian) dengan rasio pembagian yaitu 80% data *training* dan 20% data *testing*. Data pelatihan berfungsi sebagai dasar untuk mengembangkan model, sementara tujuan data pengujian digunakan untuk mengevaluasi keakuratan model yang telah dibangun [18].

5) Pembobotan kata

Tahapan pembobotan kata menerapkan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan memberikan bobot yang berbeda pada setiap kata dalam dokumen berdasarkan semakin banyak kata muncul, membuktikan seberapa umum kata tersebut dan semakin banyak dokumen mengandung sebuah *term*, maka bobot akan semakin tinggi [19]. Berikut perhitungan TF-IDF ditujukan pada persamaan 1.

$$W_{ij} = tf_{ij} \times \log\left(\frac{D}{df_j}\right) \quad (1)$$

D : Jumlah Dokumen

df_j : Jumlah Dokumen yang terdapat *term*

W_{ij} : Bobot *term* sebuah dokumen

tf_{ij} : Jumlah kemunculan *term* dalam dokumen

6) Klasifikasi *Support Vector Machine*

Algoritma klasifikasi *Support Vector Machine* yang dikembangkan oleh Vladimir Vapnik mampu memprediksi kelas berdasarkan pola yang berasal dari pembelajaran berbasis

machine learning (supervised learning). Secara sederhana, konsep SVM dapat dijelaskan sebagai usaha untuk menemukan *hyperplane optimal* yang dapat memisahkan dua kelas dalam ruang input. Dengan menghitung nilai margin dan mencari titik maksimumnya, maka akan mendapatkan *hyperplane optimal* [20]. Algoritma SVM memiliki fungsi *kernel* salah satunya yaitu *kernel linear* yang akan digunakan pada penelitian ini dan setiap *kernel* memiliki nilai parameternya sendiri. Dalam fungsi *kernel linear* digunakan parameter C. Parameter C berfungsi untuk mengontrol pertukaran antara margin dan kesalahan klasifikasi [21].

7) Evaluasi

Evaluasi dalam penelitian ini mengimplementasikan *confusion matrix* dengan prediksi *accuracy*, *precision*, dan *recall* untuk menampilkan hasil data *testing* dari uji dari model yang telah dilakukan dan membandingkan dari dua *dataset* ulasan aplikasi *mobile banking* [22]. Tabel *confusion matrix* dapat dilihat pada tabel 1.

Tabel 1 Confusion Matrix

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

- TP : Banyaknya data yang diprediksi positif dan kenyataannya positif.
- TN : Banyaknya data yang diprediksi negatif dan kenyataannya negatif.
- FP : Banyaknya data yang diprediksi positif dan kenyataannya negatif.
- FN : Banyaknya data yang diprediksi negatif dan kenyataannya positif.

8) Visualisasi

Pada tahap visualisasi, peneliti menggunakan *tools* dari *Microsoft Power BI* untuk memvisualisasikan data berupa grafik yaitu *Stacked Column Chart*, *Clustered Column Chart* dan *Wordcloud* serta penjelasan dari hasil analisis yang dilakukan. *Microsoft power BI* adalah aplikasi *Business Intelligence* yang memungkinkan pengguna untuk menampilkan informasi secara visual. Aplikasi ini sangat diperlukan karena mampu menyajikan data dalam bentuk visualisasi diagram, sehingga memudahkan para pemangku kepentingan dalam mengelola informasi dengan cepat dan kualitas yang tinggi, serta mempermudah proses pengambilan keputusan yang efektif [23].

4. Hasil dan Pembahasan

4.1. Pengumpulan Data

Scraping data ulasan aplikasi BRImo dan BCA *Mobile* pada *Play Store*, data tersebut disimpan dalam bentuk *comma-separated values (csv)*.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	userName	at	score	content											
2	Zanitha Safitri Rambe	09/04/2023 21.39	1	bca mobile ancur banget sih, verifikasi wajah gagal terus padahal identitas kumulit 092											
3	Bimbing Maryanto	09/04/2023 21.24	5	Sangat membantu dalam banyak hal.											
4	Dani Ramdani	09/04/2023 21.06	1	Sekarang ribet banget kalo ganti device ,harus verifikasi tpi gagal terus ,bca chat halo bca ,di sruh telpon layanan verifikasi lgi ,uju											
5	Abi Abdillah	09/04/2023 20.42	1	Kenapa mau verifikasi ga bisa bisa pulsa sampe abis tolong dong pihak bca di perbaiki lagi											
6	Andre Style 34	09/04/2023 20.30	4	Sya dah lama kartu ga da transaksi kira2 masih bisa ngga di gunakan lgi...soalnya mau di gunakan lgi neh											
7	Yudi Adi Purnama	09/04/2023 20.13	1	WAJIB BINTANG 1 Bagi saya SCAN wajah tidak efektif Kebanyakan ATURAN. MASIH banyak kok bank yang mempermudah akses											
8	Sri Novianti	09/04/2023 20.08	1	Kenapa uang transpar dari orang lain ke bca saya pend.tolong bantu uang saya masukin ke rek bca saya..mohon											
9	Aris Mufaroh	09/04/2023 20.08	1	Mau aktifasi m banking sudah beli pulsa 10k dua sms gagal semua habis pulsa kita itu nyari duit susah ya min kalau mau buat sis											
10	Prabuputra Exalt.purnam	09/04/2023 20.07	1	KEBANYAKAN ATURAN segala SCAN wajah. Masih banyak kok bank yang di permudak akses nya.											
11	Mr Jhon	09/04/2023 20.07	5	Tetap bertahan di mode mbanking yang disain elegan dan simple											
12	ramdani 02	09/04/2023 19.13	5	saya harap ketika selesai bikin tabungan online kartu atm bisa di kirim ke rumah seperti bank yang lain seperti line bank , bak jag											
13	Ocha Aja	09/04/2023 19.12	2	Ribet y sekrng pke foto segala											
14	Imam Mawardi	09/04/2023 18.57	2	Keyboarnya kok sering force close yah ?											
15	Teguh Sentosa	09/04/2023 18.50	5	Ok											

Gambar 2 Contoh Hasil Scraping Data Ulasan

Gambar 3 Contoh hasil *scraping* pada ulasan aplikasi BCA *Mobile*, terdapat 4 kolom yaitu *username*, *at*, *score*, *content*. Jumlah data hasil *scraping* dari masing-masing ulasan aplikasi dapat dilihat pada tabel 2.

Tabel 2 Jumlah Hasil Scraping Data	
Aplikasi	Jumlah Data
BRImo	2.000
BCA Mobile	2.000

4.2. Labelling Lexicon Based

Tahap selanjutnya yaitu pelabelan, pada tahap ini peneliti melakukan pelabelan menggunakan *Lexicon Based* dengan menggunakan program *python*. Peneliti memilah kolom yang hanya digunakan saja yaitu kolom *content* dan menambahkan kolom *score polarity* dan label untuk menyimpan data yang sudah di label menggunakan *Lexicon Based*, ulasan hanya dibagi menjadi dua sentimen yakni positif dan negatif saja, maka ulasan yang bersifat netral dan yang bukan opini akan dihapuskan. Contoh hasil pelabelan *Lexicon Based* ditunjukkan pada tabel 3.

Tabel 3 Contoh Hasil *Labelling Lexicon Based*

Content	Polarity Score	Label
Aplikasi mobile bca sangat bagus...mudah transaksi dan tidak ribet...cocok di pakai android versi 12...	7	Positif
M bca makin kesini makin jelek, masa verifikasi tidak bisa mulu dengan alasan sinyal lah apa lah,hadeuhhh...	-12	Negatif

Jumlah perbandingan data ulasan hasil pelabelan dari masing-masing aplikasi dapat dilihat pada tabel 4 dibawah ini.

Tabel 4 Jumlah Data Ulasan Hasil *Labelling*

Aplikasi	Jumlah Data	Label Positif	Label Negatif
BRimo	1.911	366	1.545
BCA Mobile	1.937	265	1.672

4.3. Preprocessing Data

1) Case Folding

Perubahan seluruh huruf/teks menjadi *lowercase*/huruf kecil. Contoh hasil dari proses *case folding* dapat dilihat pada tabel 5.

Tabel 5 Contoh Hasil dari Proses *Case Folding*

Sebelum Case Folding	Setelah Case Folding
Sistem tidak tersedia pas buat mau transfer.?? Knp itu ?? Minn	sistem tidak tersedia pas buat mau transfer.?? knp itu ?? minn

2) Cleansing

Pembersihan karakter *non alfabetis* dilakukan pada tahap ini seperti (?), (.) dan (,). Contoh hasil proses *cleansing* dapat dilihat pada tabel 6.

Tabel 6 Contoh Hasil dari Proses *Cleansing*

Sebelum Cleansing	Setelah Cleansing
sistem tidak tersedia pas buat mau transfer.?? knp itu ?? minn	sistem tidak tersedia pas buat mau transfer knp itu minn

3) Tokenizing

Pemotongan kalimat menjadi kata demi kata. Contoh hasil proses *tokenizing* dapat dilihat pada tabel 7.

Tabel 7 Contoh Hasil dari Proses *Tokenizing*

Sebelum Tokenizing	Setelah Tokenizing
sistem tidak tersedia pas buat mau transfer knp itu minn	['sistem', 'tidak', 'tersedia', 'pas', 'buat', 'mau', 'transfer', 'knp', 'itu', 'minn']

4) Word Normalization

Memperbaiki kata tidak resmi menjadi kata resmi. Contoh hasil proses *word normalization* dapat dilihat pada tabel 8.

Tabel 8 Contoh Hasil dari Proses *Word Normalization*

Sebelum Word Normalization	Setelah Word Normalization
['sistem', 'tidak', 'tersedia', 'pas', 'buat', 'mau', 'transfer', 'knp', 'itu', 'minn']	['sistem', 'tidak', 'tersedia', 'pas', 'buat', 'mau', 'transfer', 'kenapa', 'itu', 'minn']

5) *Stopword Removal*

Menghapus kata-kata yang tidak bermakna. Contoh hasil proses *stopword removal* dapat dilihat pada tabel 9.

Tabel 9 Contoh Hasil dari Proses *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
['sistem', 'tidak', 'tersedia', 'pas', 'buat', 'mau', 'transfer', 'kenapa', 'itu', 'minn']	['sistem', 'tersedia', 'transfer']

6) *Stemming*

Pemetaan suatu kata menjadi kata dasar. Contoh hasil proses *stemming* dapat dilihat pada tabel 10.

Tabel 10 Contoh Hasil dari Proses *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
['sistem', 'tersedia', 'transfer']	['sistem', 'sedia', 'transfer']

4.4. *Split Data*

Proses pemisahan dataset menjadi data *training* dan data *testing*. Pada penelitian ini implementasi split data menggunakan *python* dengan *library scikit-learn*.

```
from sklearn.model_selection import train_test_split
# Split into train and test data

train_X, test_X, train_Y, test_Y = model_selection.train_test_split(data['clean'], data['label'],
test_size=0.2, random_state=25)
```

Gambar 3 Proses *Split Data* pada *Python*

Pada gambar 4 merupakan proses *split* data menggunakan *python*, dengan pemilihan *text_size* sebesar 0.2 yaitu proposi *testing* set sebesar 20% dan *training* set 80%. Hasil pembagian pada setiap *dataset* ulasan dapat dilihat pada tabel 11 dibawah ini.

Tabel 11 Rasio Pemecehan Dataset Ulasan

Dataset Ulasan	Data <i>Training</i>	Data <i>Testing</i>
BRImo	1528	383
BCA <i>Mobile</i>	1549	388

4.5. *Pembobotan Kata*

Dalam penelitian ini, pembobotan kata menggunakan metode TF-IDF. Dilakukan menggunakan *python* menggunakan *library scikit-learn*.

```
from sklearn.feature_extraction.text import TfidfVectorizer

# TF-IDF
tfidf_vect_8020 = TfidfVectorizer(max_features=5000)
tfidf_vect_8020.fit(data['clean'].astype('U').values)
train_X_tfidf_8020 = tfidf_vect_8020.transform(df_train80['clean'].astype('U').values)
test_X_tfidf_8020 = tfidf_vect_8020.transform(df_test20['clean'].astype('U').values)

TfidfTransformer()
(0, 3534) 0.1983375624260435
(0, 3479) 0.17476807141803727
(0, 2045) 0.22280657719427482
(0, 1654) 0.5489958165105078
(3120, 3611)
```

Gambar 4 Proses TF-IDF dengan *Python*

Gambar 5 adalah proses TF-IDF pada *python* hasil dari proses TF-IDF menghasilkan transformasi data dengan jumlah matriks 3.120 x 3.611 atau 3.120 dokumen dan 3.611 term.

4.6. *Klasifikasi Support Vector Machine*

Pada tahap pemodelan ini peneliti menggunakan fungsi kernel yang disediakan algoritma SVM yaitu, *kernel linear*. Dalam fungsi *kernel linear* terdapat ukuran yang digunakan yaitu parameter C.

```
# lin_svm=svm.LinearSVC()
lin_svm = svm.SVC(kernel='linear', C=1)
lin_svm.fit(train_X_tfidf_8020, df_train80['label'])
#lin_svm.fit(test_vectors, data_test['sentiment'])
y_train_hat = lin_svm.predict(train_X_tfidf_8020)
y_test_hat = lin_svm.predict(test_X_tfidf_8020)

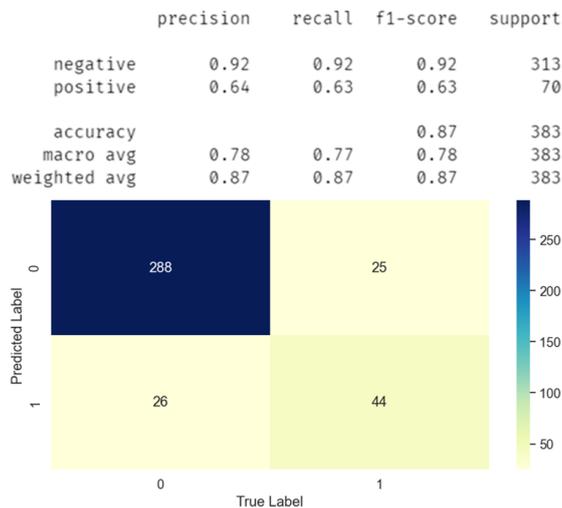
print("train accuracy= ", np.mean(y_train_hat == df_train80['label'])*100)
print("test accuracy= ", np.mean(y_test_hat == df_test20['label'])*100)
```

Gambar 5 Proses Klasifikasi SVM

Gambar 6 merupakan proses pada saat melakukan klasifikasi SVM menggunakan Python dengan bantuan library scikit-learn. Pada penelitian ini nilai parameter C yang diberikan yaitu 10 sebagai analisa mencari terbaik pada klasifikasi sentimen aplikasi BRImo dan BCA mobile.

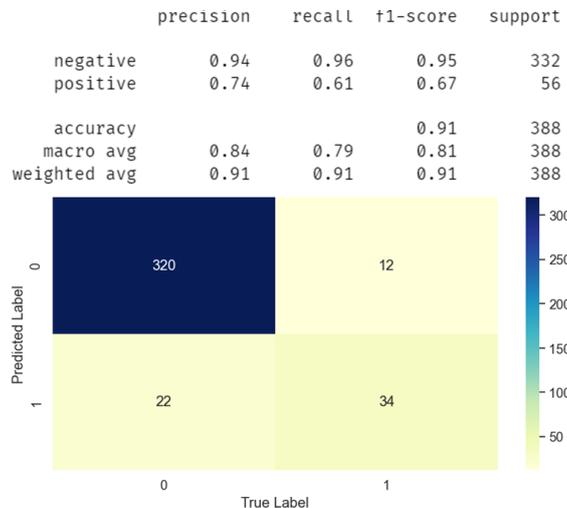
4.7. Evaluasi

Pada hasil klasifikasi yang telah dijalankan sebelumnya, untuk menganalisa dan menghitung kinerja pada algoritma SVM, peneliti menggunakan confusion matrix untuk menghitung accuracy, precision dan recall menggunakan python dengan library scikit-learn.



Gambar 6 Confusion Matrix Data Uji BRImo

Gambar 7 confusion matrix aplikasi BRImo pengujian yang diperoleh dengan menggunakan SVM menunjukkan accuracy sebesar 87%, dimana nilai precision 64% dan recall 63%.



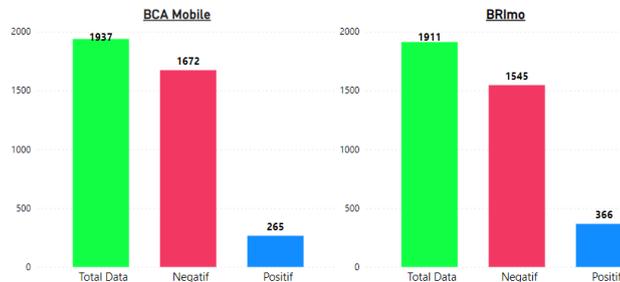
Gambar 7 Confusion Matrix Data Uji BCA Mobile

Gambar 8 *confusion matrix* aplikasi BCA Mobile pengujian yang diperoleh dengan menggunakan SVM menunjukkan *accuracy* sebesar 91%, dimana nilai *precision* 74% dan *recall* 67%.

4.8. Visualisasi

Pada tahap ini dilakukan visualisasi dengan menampilkan grafik yaitu *Stacked Column Chart*, *Clustered Column Chart* dan *Wordcloud* menggunakan *Microsoft Power BI*.

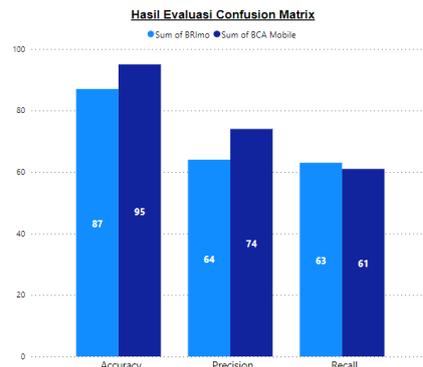
a) *Stacked Column Chart*



Gambar 8 Diagram *Labelling* BRImo dan BCA Mobile

Gambar 9 memvisualisasikan Data Ulasan aplikasi BRImo berdasarkan label, dari total jumlah ulasan sebanyak 1.911 terdapat label positif berjumlah 366 ulasan atau 19,15%, dan label negatif berjumlah 1.545 ulasan atau 80,84%. Data Ulasan dari aplikasi BCA Mobile berdasarkan label, dari total jumlah ulasan sebanyak 1.937 terdapat label positif berjumlah 265 ulasan atau 13,68%, dan label negatif berjumlah 1.672 ulasan atau 86,31%.

b) *Clustered Column Chart*



Gambar 9 Diagram Penilaian *Confusion Matrix*

Gambar 10 menunjukkan hasil perbandingan evaluasi *confusion matrix* pada aplikasi BRImo dan BCA Mobile bahwasannya didapatkan perhitungan *accuracy*, *precision* dan *recall*.

c) *WordCloud*



Gambar 11 *WordCloud* BRImo



Gambar 10 *WordCloud* BCA Mobile

Gambar 11 & 12 hasil visualisasi *WordCloud* menampilkan sebanyak 150 kata teratas dari ulasan BRImo & BCA *Mobile* baik itu yang mengandung sentimen positif maupun sentimen negatif.

Berdasarkan rangkaian proses yang telah dilakukan pada penelitian ini seperti yang tersemat pada gambar 9 hasil proses *labelling* menggunakan *Lexicon Based* pada ulasan setiap aplikasi bahwa aplikasi BRImo mendapatkan sentimen positif berjumlah 366 atau 19,15% dan sentimen negatif berjumlah 1.545 atau 80,84%. Sedangkan untuk aplikasi BCA *Mobile* mendapatkan sentimen positif berjumlah 265 atau 13,68% dan sentimen negatif berjumlah 1.672 atau 86,31%. Dari hasil proses *labelling* tersebut lalu dilakukan klasifikasi menggunakan algoritma *Support Vector Machine* dengan pengujian evaluasi model menggunakan *confusion matrix* seperti yang tersemat pada gambar 10 Hasil Pengujian *Confusion Matrix* pada aplikasi BRImo memperoleh *accuracy* 87%, *precision* 64%, *recall* 63%. Dan pada aplikasi BCA *Mobile* didapatkan *accuracy* 91%, *precision* 74%, *recall* 61%.

Dari hasil penelitian ini merupakan penelitian dengan menggunakan keterbaharuan data dan menggunakan algoritma *Support Vector Machine* yang menjadi perbedaan dengan penelitian sebelumnya yaitu pada penelitian yang dilakukan oleh Khoiril Insan et al., (2023) [9] dimana penelitian analisis sentimen ulasan BRImo dengan *Naïve Bayes* proses *labelling* dilakukan secara manual didapat sentimen positif lebih banyak sebesar 1012 atau 53,09% dibandingkan sentimen negatif sebesar 894 atau 46,90% serta memperoleh *accuracy* 84,52%. Lalu pada penelitian sebelumnya yaitu Sari et al., (2023) [10] pada penelitian analisis sentimen ulasan BCA *Mobile* menggunakan *Naïve Bayes* dengan penentuan label data berdasarkan *rating* diperoleh 250 data ulasan sentimen positif dan sentimen negatif sebanyak 250 data, hasil proses klasifikasi menggunakan *Naïve Bayes* mendapatkan *accuracy* 82%. Dari kaitan yang di dapat dengan dari penelitian terdahulu bahwa pada penelitian ini dengan jumlah data yang lebih banyak serta variatif dengan proses *labelling* menggunakan *Lexicon Based* mampu mengoptimalkan proses klasifikasi dengan menggunakan algoritma *Support Vector Machine* dapat disimpulkan bahwa algoritma *Support Vector Machine* memiliki performa yang baik didalam melakukan klasifikasi ulasan Aplikasi *Mobile Banking* (BRImo dan BCA *Mobile*) di *Play Store*.

5. Simpulan

Berdasarkan hasil penelitian analisis sentimen yang telah dilakukan menggunakan algoritma *Support Vector Machine*. Proses analisis dengan awali *scraping* ulasan BRImo dan BCA *Mobile* menggunakan *python*, lalu masuk ke tahap *preprocessing* (*case folding, cleansing, tokenizing, word normalization, stopword, dan stemming*). Data dilakukan *labelling* secara manual, setelah itu masuk pada tahap *split* data dengan rasio 80% data *training* dan 20% data *testing*, lalu tahap pembobotan kata menggunakan TF-IDF, dan hasil pembobotan kata dilakukan *modelling* dengan *Support Vector Machine* dan evaluasi dari model menggunakan *confusion matrix* menggunakan pengukuran *accuracy, precision* dan *recall*.

Klasifikasi sentimen ulasan aplikasi BRImo dan BCA *Mobile* dalam pengujian menggunakan *Support Vector Machine* pada aplikasi BRImo didapatkan *accuracy* 94%, *precision positive* 96%, *precision negative* 93%, *recall positive* 85%, dan *recall negative* 98%. Lalu pada aplikasi BCA *Mobile* *accuracy* 95%, *precision positive* 96%, *precision negative* 94%, *recall positive* 72%, dan *recall negative* 99%. Diketahui bahwa SVM memiliki performa yang baik dalam penelitian ini. Berdasarkan hasil tersebut diketahui bahwa aplikasi BCA *Mobile* memiliki sentimen negatif yang lebih banyak dibandingkan aplikasi BRImo bahwa pengguna BCA *Mobile* merasa tidak puas terhadap aplikasi tersebut. Pada penelitian berikutnya dapat ditambahkan menggunakan *kernel non linier* pada SVM seperti *polynomial, radial basis function* dan *sigmoid* atau menambah/menggunakan algoritma yang lain seperti *random forest, K-NN, regresi linear* atau algoritma *machine learning* yang lainnya.

Daftar Referensi

- [1] F. Bei and S. Saepudin, "Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm)," *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [2] M. Hamka, N. Alfatari, and D. Ratna Sari, "Analisis Sentimen Produk Kecantikan Jenis Serum Menggunakan Algoritma Naïve Bayes Classifier," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 64, 2022, doi: 10.30865/json.v4i1.4740.

- [3] A. Putri and A. Muzakir, "Analisis Sentimen Cyberbullying Kpop Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Ilm. Indones.*, vol. 2, no. 8.5.2017, pp. 2003–2005, 2022, [Online]. Available: <https://jurnal.syntaxliterate.co.id/index.php/syntaxliterate/article/view/9334>
- [4] A. Deolika, K. Kusri, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [5] F. Romadoni, Y. Umaidah, and B. N. Sari, "Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 2, pp. 247–253, 2020, doi: 10.32736/sisfokom.v9i2.903.
- [6] A. F. Riyadi, F. R. Rahman, M. A. Nofa Pratama, M. K. Khafidli, and H. Patria, "Pengukuran Sentimen Sosial Terhadap Teknologi Kendaraan Listrik: Bukti Empiris di Indonesia," *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 11, no. 2, p. 141, 2021, doi: 10.36448/expert.v11i2.2171.
- [7] H. P. Doloksaribu and Y. T. Samuel, "Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 16, no. 1, pp. 1–11, 2022, doi: 10.47111/jti.v16i1.3747.
- [8] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naive Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [9] M. Khoirul Insan, U. Hayati, and O. Nurdiawan, "ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023.
- [10] W. F. Sari, R. Rahim, and F. Adrianto, "Analisis Sentiment Review Pengguna Bca Mobile Menggunakan Teks Mining," *CAKRAWALA Repos. IMWI*, vol. 6, no. 2, pp. 981–987, 2023, doi: <https://doi.org/10.52851/cakrawala.v6i2.295>.
- [11] N. E. Oktaviana, Y. A. Sari, and I. Indriati, "Analisis Sentimen terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features dan Support Vector Machine," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, p. 357, 2022, doi: 10.25126/jtiik.2022925625.
- [12] R. D. Wahyuni and A. N. Utomo, "Penggunaan Metode Lexicon Untuk Analisis Sentimen Pada Ulasan Aplikasi Kai Access Di Google Play Store," *J. Rekayasa Inf.*, vol. 11, no. 2, pp. 134–135, 2022, [Online]. Available: <https://ejournal.istn.ac.id/index.php/rekayasainformasi/article/view/1415>
- [13] W. M. P. Dhuhita, M. F. K. A. Darmawan, L. Triana, and N. Ankisqiantari, "Perbandingan Algoritma Supervised Learning untuk Klasifikasi Judul Skripsi Berdasarkan Bidang Dosen," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, pp. 427–437, 2022, doi: 10.28932/jutisi.v8i2.4960.
- [14] I. B. N. W. Manuaba, G. R. Dantes, and G. Indrawan, "Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine Dengan Penambahan Algoritma Levenshtein Distance," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 5, no. 2, pp. 9–17, 2022, doi: 10.47970/siskom-kb.v5i2.261.
- [15] P. Aditiya, U. Enri, and I. Maulana, "Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1020, 2022, doi: 10.30865/jurikom.v9i4.4673.
- [16] N. Aliyah Salsabila, Y. Ardhitto Winatmoko, A. Akbar Septiandri, and A. Jamal, "Colloquial Indonesian Lexicon," *Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018*, pp. 226–229, 2019, doi: 10.1109/IALP.2018.8629151.
- [17] Andreyestha and Q. Nur Azizah, "Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 108–116, 2022, doi: 10.29408/jit.v5i1.4581.
- [18] D. Pajri, Y. Umaidah, and T. N. Padilah, "K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 242–253, 2020, doi: <https://doi.org/10.28932/jutisi.v6i2.2658>.
- [19] S. Y. Pangestu, Y. Astuti, and L. D. Farida, "Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia," *J. Mantik Penusa*, vol. 3, no.

- 1, pp. 236–241, 2019, [Online]. Available: <https://t.co/eF>
- [20] G. R. Ditami, E. F. Ripanti, and H. Sujaini, “Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 3, pp. 508–516, 2022.
- [21] Styawati, Andi Nurkholis, Zaenal Abidin, and Heni Sulistiani, “Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 904–910, 2021, doi: 10.29207/resti.v5i5.3380.
- [22] F. E. Kavabilla, T. Widiharih, and B. Warsito, “Analisis Sentimen pada Ulasan Aplikasi Investasi Online Ajaib pada Google Play Menggunakan Metode Support Vector Machine dan Maximum Entropy,” *J. Gaussian*, vol. 11, no. 4, pp. 542–553, 2023, doi: 10.14710/j.gauss.11.4.542-553.
- [23] J. E. Bororing and A. Pasadi, “Implementasi Microsoft Power Bi Untuk Dashboard Visualisasi Data Akademik Mahasiswa Fakultas Teknik Universitas Janabadra,” *J. Inf. Interaktif*, vol. 7, no. 2, pp. 149–155, 2022.