Journal of Electrical Engineering and Computer (JEECOM)
Vol. 6, No. 2 (2024), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Analisis Sentimen Terhadap Ulasan Aplikasi Shopee di Google Play Store Menggunakan Metode TF-IDF dan Long Short-Term Memory (LSTM)

Musfiroh[1], Abu Tholib[2], Zainal Arifin[3]
[1,2,3] Teknik Informatika, Universitas Nurul Jadid, Probolinggo, Indonesia

Article Info

Article history:

Diterima 27 Mey 2024
Revisi 23 Juli 2024
Diterbitkan 4 Oktober 2024

Keywords:

Analisis Sentimen
LSTM
Shopee
Text Mining
TF-IDF

ABSTRAK

Pengunjung Shopee semakin meningkat dari tahun 2022 hingga 2023. Karena peningkatan itu, semakin banyak pengguna yang berkomentar negatif atau positif. Maka, mengetahui sentimen pengguna pada aplikasi *Shopee* dapat mengetahui perilaku pelanggan dan meningkatkan penjualan. Penelitian ini menggunakan metode *TF-IDF* dan *algoritma LSTM*. Adapun tahapan penelitian seperti *scrapping data* yang menggunakan ulasan pengguna aplikasi *Shopee* di *Google Play Store* sebanyak 3565 data. Lalu data dikategorikan menjadi tiga kelas: positif, netral, dan negatif. Proses *preprocessing* meliputi *Tokenization*, *Normalization*, *Stopword*, dan *Stemming.* Selanjutnya dilakukan proses *train data* dan *data test* sebesar 8:2. Lalu melakukan vektorisasi dengan *TF-IDF*, melatih model dengan penggabungan *TF-IDF* dan *LSTM (Long Short-Term Memory)*, serta menggunakan metrics untuk mengevaluasi model dan visualisasi menggunakan *word cloud.* menghasilkan akurasi sebesar 83% dengan nilai *loss* (kerugian) sebesar 0.1385. Model memiliki kemampuan cukup baik dalam memprediksi kelas negatif dan positif tetapi kurang efektif untuk kelas netral karena data yang kurang seimbang.

Corresponding Author:
Abu Tholib,
Universitas Nurul Jadid, Karanganyar Paiton, Probolinggo 67291, Indonesia
Email: ebuenje@gmail.com

1. PENDAHULUAN

Shopee saat ini menjadi salah satu *marketplace* yang paling diminati masyarakat Indonesia pada tahun 2023 dengan total pengunjung pada kuartal 1 sebanyak 157 juta dan Tokopedia sebanyak 117 juta[1]. Pada tahun sebelumnya, *Shopee* menempati posisi kedua pada kuartal 1 dengan total pengunjung terbanyak sebanyak 132 juta setelah Tokopedia yang berjumlah 157 Juta [2]. Semakin meningkatnya pengunjung *Shopee* dari tahun ke tahun, maka akan semakin banyak komentar dari pengguna baik itu komentar positif atau negatif. Maka, analisis sentimen diperlukan untuk meningkatkan layanan *Shopee* serta pengambilan keputusan.

Gambar 1. Pengunjung *E-Commerce* terbanyak tahun 2022 dan 2023 (Kuartal 1)

ISSN: 2715-6427

Dalam penelitian analisis sentimen yang telah dilakukan, beberapa penelitian menggunakan metode seperti *Support Vector Machine (SVM*) [3], *Stochastic Gradient Descent* [4] dan *Naive Bayes* [5]. Metode *Support Vector Machine (SVM)* memiliki kelebihan dalam klasifikasi teks yang tidak linier dan *opinion mining* dibandingkan dengan beberapa metode *machine learning* lainnya seperti *Naïve Bayes, Decision Tree*, dan lainnya, namun masih memiliki kekurangan dalam menangani permasalahan data yang berdimensi tinggi secara efisien [3]. Algoritma *Stochastic Gradient Descent* memiliki kelebihan seperti mudah diimplementasikan, tidak membutuhkan memori yang besar untuk menyimpan data, dan cocok untuk data *real-time* karena dapat diupdate secara terus menerus. Namun memiliki kekurangan dalam kestabilan perubahan data, dan hasil yang tidak konsisten karena kurangnya data [6]. *Naive Bayes Classifier* memiliki kelebihan dalam kemudahan implementasi, kinerja yang baik dan cepat dalam memproses data besar, optimal dalam melakukan klasifikasi, serta kemampuan untuk mengatasi masalah ketidakseimbangan dalam kelas. Namun, metode ini memiliki kelemahan seperti sensitivitas terhadap kualitas data, dan kinerja yang dapat dipengaruhi oleh representasi teks dan ketidakseimbangan kelas [5], [7].

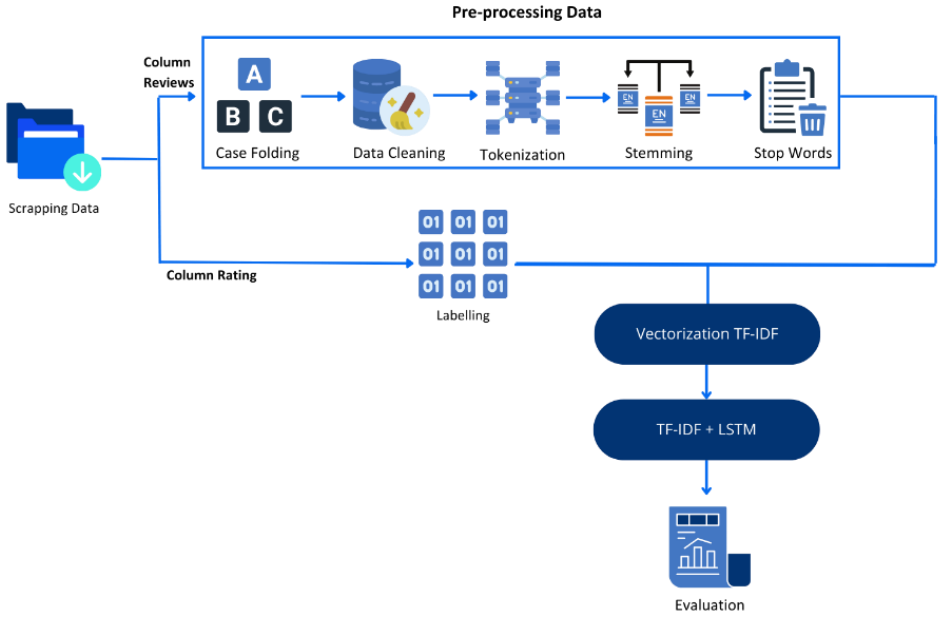
Penelitian ini menggunakan algoritma *LSTM (Long Short-Term Memory)*. Algoritma ini memiliki kemampuan untuk mengingat informasi dalam teks untuk jangka waktu yang lama *(long term dependency)*, yang memungkinkannya untuk lebih baik dalam memahami konteks dan emosi dalam teks [8]. Kelebihan ini memungkinkan LSTM untuk membuat prediksi yang lebih akurat. Lalu, *LSTM (Long Short-Term Memory)* memiliki sel memori yang dapat mengatasi masalah hilangnya informasi yang sering terjadi pada model jaringan saraf lainnya. Namun, *LSTM (Long Short-Term Memory)* rentan terhadap *overfitting*, terutama jika dataset tidak seimbang atau ukuran dataset kecil [9].

TF-IDF (Term Frequency-Inverse Document Frequency) membantu mengatasi *kelemahan LSTM (Long Short-Term Memory)* dalam mempertimbangkan bobot kata-kata dalam teks. Meskipun *LSTM (Long Short-Term Memory)* tidak secara langsung memperhitungkan tingkat penting kata, *TF-IDF (Term Frequency-Inverse Document Frequency)* memberikan bobot yang lebih tinggi pada kata-kata yang penting. Dengan demikian, penggunaan *TF-IDF (Term Frequency-Inverse Document Frequency)* membantu *LSTM (Long Short-Term Memory)* dalam memahami konteks dan makna kata dalam teks secara lebih efektif [10].

Penelitian ini akan menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)* dan algoritma *LSTM (Long Short-Term Memory)* dengan menggunakan *data review* aplikasi *Shopee* dari *Google Play Store* menggunakan teknik *scrapping* dengan *Python*. Data kemudian diklasifikasikan menjadi 3 kategori, yakni positif, negatif, dan netral. Penelitian bertujuan untuk memahami sentimen pengguna sebagai pertimbangan bagi pengembang untuk meningkatkan layanan aplikasi.

2. METODE

Metode penelitian dalam analisis sentimen terhadap ulasan aplikasi shopee di *Google Play Store* memiliki beberapa tahapan seperti yang ditunjukkan pada gambar 2.



Gambar 2. Alur Penelitian

2.1. Pengumpulan Data

Penelitian ini menggunakan data ulasan dari aplikasi Shopee di *Google Play Store*. Pengambilan data menggunakan teknik *scrapping* dengan *Google Colaboratory*. Data yang digunakan dalam analisis ini hanya kolom rating dan *review.*

2.2. *Pre-processing* Data

Setelah mengumpulkan data mentah, data *review* kemudian masuk ke tahapan *pre-processing data* seperti dibersihkan, disederhanakan, dan ditransformasi agar mudah diolah dan meningkatkan kualitas data [11], [12]. *Preprocessing data* yang digunakan meliputi *Tokenization*, *Normalization*, *Stopword*, dan *Stemming* [13]. Tahapan *pre-processing* dilakukan secara terpisah antara data *review* dan data *score*/rating.

2.2.1. *Labelling*

Data hasil *scrapping* kemudian diberikan label atau kategori pada data untuk mengidentifikasi informasi. Data yang akan diberikan label adalah kolom rating. Penelitian ini akan melabeli dataset menjadi 3 kategori. Untuk rating 1 dan 2 akan diberi label "negatif", rating 3 diberi label "netral", dan rating 4 dan 5 akan diberi label "positif" [14].

2.2.2. *Case Folding*

Setelah data diberi label, kemudian huruf kapital pada data *review* akan diubah menjadi huruf kecil atau *Case folding*. Proses ini menggunakan fungsi *"lower()"* dari *Natural Language Toolkit (NLTK)* dalam bahasa pemrograman *Python* [15]. Lalu, data akan menyisakan huruf saja dengan menghapus tanda baca, emoji, dan angka [16].

2.2.3. *Data Cleaning*

Selanjutnya, data yang berbentuk kata singkatan, pemanjangan kata, dan *slang*, akan diubah menjadi kata aslinya agar mudah dipahami [16], [17]. Selain itu, data yang hanya berisikan rating tanpa ada review *(missing value)* akan dihapus agar tidak mempengaruhi kinerja model [18].

2.2.4. *Tokenization*

Setelah melakukan *data cleaning*, data akan dipecah menjadi token atau satuan teks. Proses ini menggunakan *Natural Language Toolkit (NLTK)* untuk tokenisasi. Proses ini akan membagi tokenisasi per kata [19], [20].

2.2.5. *Stemming*

Setelah melakukan tokenisasi, data masuk kedalam tahapan *stemming*. Data yang memiliki imbuhan didalamnya akan diubah menjadi bentuk kata dasarnya. Proses ini menggunakan *library* sastrawi khusus untuk Bahasa Indonesia. [21].

2.2.6. *Stop Words*

Setelah melakukan *stemming*, proses terakhir dalam *preprocessing* yaitu penghapusan kata *stop words* pada data yang dilakukan dengan mengidentifikasi kata-kata tidak informatif dan umum dalam teks. Proses ini menggunakan *library python Natural Language Toolkit (NLTK)* untuk Bahasa Indonesia. Kata yang dihapus biasanya seperti kata depan, kata sambung, dan kata bantu. Selanjutnya, menghapus kata yang tidak relevan seperti beberapa kata yang memiliki sentimen negatif, namun masuk ke dalam kategori positif maupun sebaliknya [22].

2.3. *TF-IDF (Term Frequency-Inverse Document Frequency)*

Setelah data dibersihkan, data akan dibagi menjadi *data test*, dan *data train* [23]. Setelah itu, data yang sudah dibagi diberikan bobot sesuai dengan banyaknya kemunculan kata dan seberapa penting arti kata tersebut menggunakan *library python* dari *scikit-learn* yaitu *TfidfVectorizer*[24]. Lalu data yang semula berbentuk *string* akan diubah menjadi vektor fitur sesuai dengan bobot kata.

$$w_{i,j} = tf_{i,j} \times log\ log\ \left(\frac{N}{df_i}\right) \quad (1)$$

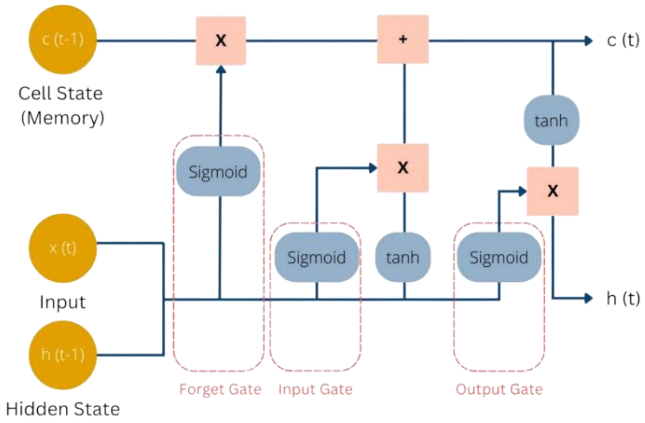
Perhitungan *TF-IDF (Term Frequency-Inverse Document Frequency)* dalam pembobotan kata di mana $w_{i,j}$ merupakan nilai bobot yang diberikan pada istilah (t_j) dalam dokumen (d_i). Sementara $tf_{i,j}$ adalah jumlah kemunculan istilah (t_j) dalam dokumen (d_i). N adalah total jumlah dokumen dalam basis data dan df_j adalah

ISSN: 2715-6427

jumlah dokumen yang mengandung istilah (t_j) (setidaknya satu kata merupakan istilah (t_j)). Meskipun nilai $tf_{i,j}$bervariasi, jika $N = df_j$, maka hasilnya akan menjadi 0 (nol) karena hasil dari log 1 dalam perhitungan IDF [25]. Setelah proses pembobotan kata, label yang awalnya berbentuk huruf akan diubah menjadi numerik dengan menggunakan *label encoder* dari *Sklearn* [26].

2.4. *LSTM (Long Short-Term Memory)*

Data yang telah diberikan bobot kemudian dimasukkan ke dalam model *Long Short-Term Memory (LSTM)*. Dalam proses ini akan mendefinisikan model *Long Short-Term Memory (LSTM)* yang dibuat seperti Gerbang Lupa *(Forget Gate)*, Gerbang Input *(Input Gate)*, dan Gerbang Keluaran *(Output Gate)* yang kemudian menghasilkan output [26], [27]. Kemudian, menggabungkan *TF-IDF (Term Frequency-Inverse Document Frequency)* dan *Long Short-Term Memory (LSTM)* dengan *concatenate* dari *TensorFlow* [28] Dan terakhir melatih model yang telah dibuat.



Gambar 3. Arsitektur LSTM

2.5. Evaluasi Hasil

Setelah melatih data ke dalam model, hasil analisis tersebut kemudian akan dievaluasi untuk menghitung tingkat akurasi menggunakan metriks evaluasi dari *Sklearn* dan mengevaluasi kinerja model[29]. Setelah itu, langkah terakhir dalam evaluasi adalah menampilkan *Word Cloud* untuk setiap kategori yang terdiri dari kata-kata yang sering muncul dalam analisis [30].

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data scrapping menggunakan *library python* yaitu *google_play_scraper*. Ulasan didefinisikan berdasarkan argumen *"lang", "country",* dan *"count". "lang = 'id'"* mengatur bahasa ulasan menjadi bahasa Indonesia, dan "*country = 'id'"* menetapkan negara sebagai Indonesia. *"count = 3000",* untuk meminta data terbaru sebanyak 3000 ulasan dari aplikasi. Dari data yang ingin dilakukan *scrapping*, variabel yang diambil meliputi username, tanggal, komentar, dan rating. Selanjutnya dilakukan filtrasi untuk mengambil ulasan yang diberikan dalam rentang waktu September 2023 hingga Mei 2024. Setelah proses filtrasi, didapatkan hasil sebanyak 3565 data ulasan yang memenuhi kriteria tersebut. Hasil dari *scrapping* terdapat pada tabel dibawah ini.

Tabel 1. Data Hasil *Scrapping*

TABLE I. DATA HASIL *SCRAPPING*

Username	**Tanggal**	**Rating**	**Komentar**
K*****a B******** *****s	5/13/2024 8:42:54	5	Sudah bertahun-tahun saya belanja di Shopee, sampai sekarang pelayanannya bagus sih menurut aq, baik dari tokonya,juga kurir nya. sejauh ini g pernah terjadi sesuatu yg mengecewakan sih, untuk saya sebagai pelanggan
L**a R**i	12/31/2023 12:35:12	1	Akhir ini shopee kenapa ya? Kesel banget sama pengirimannya, kurir ga ada ngasih kabar mau ngantar tapi kok gagal mulu dan akhirnya dibatalin dan dikembalikan kepenjual.

U*S I***S	12/10/2023 11:42:37	3	Update terbaru,malah down..server error..jdi bagaimana kita mau belanja..? Mau masuk ajah susah banget.
I**a Z*******a	12/25/2023 13:51:34	2	Terlalu banyak biaya yang lain lain nya, di tambah lagi ongkir nya mahal banget sekarang harga barang ga seberapa ongkir sama biaya yang lain lain nya gede banget
S*****g N***a	5/13/2024 10:41:40	4	saya pengguna baru dari bulan kemarin. jika dapat penjual yang amanah dan terpercaya pasti kita happy, 5 kali belanja ada 2 kali penjual yang tidak amanah, barangnya gak seperti deskripsi dari penjual pas sampai beda.

3.2. *Pre-processing Data*

Setelah *scrapping data*, kemudian data tersebut masuk ke tahap *preprocessing data*. Tahap ini ada beberapa proses yang dilakukan seperti memberikan label sesuai dengan rating yang diberikan, *case folding, stop words, stemming,* dan *data cleaning*.

3.2.1. *Labelling*

Sebelum memberikan label pada data, beberapa kolom dihapus dan hanya menyisakan kolom rating dan komentar. Kemudian, data tersebut disusun dengan memberikan label positif, netral, dan negatif. Dari total 3565 data, terdapat 1620 data dengan rating 1, 483 data dengan rating 2, 433 data dengan rating 3, 295 data dengan rating 4, dan 734 data dengan rating 5. Setelah memberikan label sesuai dengan rating, data tersebut terbagi menjadi 2103 data dengan label negatif, 433 data dengan label netral, dan 1029 data dengan label positif. Berikut beberapa hasil dari *labelling*.

Tabel 2. *Labelling Data*

TABLE II. *LABELLING DATA*

Komentar	Rating	Label
Sudah bertahun-tahun saya belanja di Shopee, sampai sekarang pelayanannya bagus sih menurut aq, baik dari tokonya,juga kurir nya. sejauh ini g pernah terjadi sesuatu yg mengecewakan sih, untuk saya sebagai pelanggan	5	positif
Akhir ini shopee kenapa ya? Kesel banget sama pengirimannya, kurir ga ada ngasih kabar mau ngantar tapi kok gagal mulu dan akhirnya dibatalin dan dikembalikan kepenjual.	1	negatif
Update terbaru,malah down..server error..jdi bagaimana kita mau belanja..? Mau masuk ajah susah banget.	3	netral
Terlalu banyak biaya yang lain lain nya, di tambah lagi ongkir nya mahal banget sekarang harga barang ga seberapa ongkir sama biaya yang lain lain nya gede banget	2	negatif
saya pengguna baru dari bulan kemarin. jika dapat penjual yang amanah dan terpercaya pasti kita happy, 5 kali belanja ada 2 kali penjual yang tidak amanah, barangnya gak seperti deskripsi dari penjual pas sampai beda.	4	positif

3.2.2. *Case Folding*

Kemudian, data yang memiliki huruf besar, tanda baca, dan angka dalam setiap data akan dilakukan *case folding*. Langkah awal yakni mengubah huruf besar menjadi huruf kecil. Kemudian menghapus tanda baca yang terdapat dalam data. Dan yang terakhir adalah menghapus angka. Hasil dari *case folding* tertera pada tabel dibawah ini.

Tabel 3. *Case Folding*

TABLE III. *CASE FOLDING*

Case Folding	Label
sudah bertahun tahun saya belanja di shopee sampai sekarang pelayanannya bagus sih menurut aq baik dari tokonya juga kurir nya sejauh ini g pernah terjadi sesuatu yg mengecewakan sih untuk saya sebagai pelanggan	positif
akhir ini shopee kenapa ya kesel banget sama pengirimannya kurir ga ada ngasih kabar mau ngantar tapi kok gagal mulu dan akhirnya dibatalin dan dikembalikan kepenjual	negatif
update terbaru malah down server error jdi bagaimana kita mau belanja mau masuk ajah susah banget	netral

ISSN: 2715-6427

terlalu banyak biaya yang lain lain nya di tambah lagi ongkir nya mahal banget sekarang harga barang ga seberapa ongkir sama biaya yang lain lain nya gede banget	negatif
saya pengguna baru dari bulan kemarin jika dapat penjual yang amanah dan terpercaya pasti kita happy kali belanja ada kali penjual yang tidak amanah barangnya gak seperti deskripsi dari penjual pas sampai beda	positif

3.2.3. *Data Cleaning*

Tahap selanjutnya adalah *data cleaning* untuk merapihkan data yang memiliki makna dengan mendefinisikan arti dari kata-kata yang ditulis secara singkat oleh para pengguna ke dalam arti sesungguhnya seperti kata *"mslh"* menjadi *"masalah"*, dan "*bgs*" menjadi "*bagus*". Lalu mengubah kata *slang* seperti "*mager*" menjadi "*malas gerak*", dan "*gercep*" menjadi "*gerak cepat*". Kemudian menghapus kata yang memiliki pemanjangan kata seperti "*iyaaaaa*", "*terbaruuuu*", dan" *lamaaaa*". Selain itu, dalam tahap ini juga mengubah kata tidak baku menjadi kata baku. Berikut hasil *data cleaning* seperti yang ditampilkan pada tabel.

Tabel 4. *Data Cleaning*

TABLE IV. *DATA CLEANING*

Data Cleaning	**Label**
sudah bertahun tahun saya belanja di shopee sampai sekarang pelayanannya bagus sih menurut aq baik dari tokonya juga kurir nya sejauh ini g pernah terjadi sesuatu yg mengecewakan sih untuk saya sebagai pelanggan	positif
akhir ini shopee kenapa ya kesal banget sama pengirimannya kurir ga ada memberi kabar mau mengantar tapi kok gagal terus dan akhirnya dibatalkan dan dikembalikan kepenjual	negatif
update terbaru malah down server error jadi bagaimana kita mau belanja mau masuk saja susah banget	netral
terlalu banyak biaya yang lain lain nya di tambah lagi ongkos kirim nya mahal banget sekarang harga barang ga seberapa ongkir sama biaya yang lain lain nya besar banget	negatif
saya pengguna baru dari bulan kemarin jika dapat penjual yang amanah dan terpercaya pasti kita happy kali belanja ada kali penjual yang tidak amanah barangnya gak seperti deskripsi dari penjual pas sampai beda	positif

3.2.4. *Tokenization*

Selanjutnya, kata yang telah dibersihkan akan masuk kedalam tahap tokenisasi. Dalam data ini, tokenisasi menggunakan *word_tokenize* dari *Natural Language Toolkit (NLTK).* Data akan dipecah menjadi satuan kata seperti contoh kalimat *"update terbaru malah down server error jadi bagaimana kita mau belanja mau masuk saja susah banget"* menjadi "update", "terbaru", "malah", "down", "server", "error, "jadi", "bagaimana", "kita", "mau", "belanja", "mau", "masuk", "saja", "susah", "banget". Berikut hasil data yang telah ditokenisasi dalam tabel berikut.

Tabel 5. *Tokenization*

TABLE V. *TOKENIZATION*

Data Cleaning	***Tokenization***	**Label**
sudah bertahun tahun saya belanja di shopee sampai sekarang pelayanannya bagus sih menurut aq baik dari tokonya juga kurir nya sejauh ini g pernah terjadi sesuatu yg mengecewakan sih untuk saya sebagai pelanggan	'sudah', 'bertahun', 'tahun', 'saya', 'belanja', 'di', 'shopee', 'sampai', 'sekarang', 'pelayanannya', 'bagus', 'sih', 'menurut', 'aq', 'baik', 'dari', 'tokonya', 'juga', 'kurir', 'nya', 'sejauh', 'ini', 'g', 'pernah', 'terjadi', 'sesuatu', 'yg', 'mengecewakan', 'sih', 'untuk', 'saya', 'sebagai', 'pelanggan'	positif
akhir ini shopee kenapa ya kesal banget sama pengirimannya kurir ga ada memberi kabar mau mengantar tapi kok gagal terus dan akhirnya dibatalkan dan dikembalikan kepenjual	'akhir', 'ini', 'shopee', 'kenapa', 'ya', 'kesal', 'banget', 'sama', 'pengirimannya', 'kurir', 'ga', 'ada', 'memberi', 'kabar', 'mau', 'mengantar', 'tapi', 'kok', 'gagal', 'terus', 'dan', 'akhirnya', 'dibatalkan', 'dan', 'dikembalikan', 'kepenjual'	negatif

update terbaru malah down server error jadi bagaimana kita mau belanja mau masuk saja susah banget	'update', 'terbaru', 'malah', 'down', 'server', 'error', 'jadi', 'bagaimana', 'kita', 'mau', 'belanja', 'mau', 'masuk', 'saja', 'susah', 'banget'	netral
terlalu banyak biaya yang lain lain nya di tambah lagi ongkos kirim nya mahal banget sekarang harga barang ga seberapa ongkir sama biaya yang lain lain nya besar banget	'terlalu', 'banyak', 'biaya', 'yang', 'lain', 'lain', 'nya', 'di', 'tambah', 'lagi', ongkos, 'kirim', 'nya', 'mahal', 'banget', 'sekarang', 'harga', 'barang', 'ga', 'seberapa', 'ongkir', 'sama', 'biaya', 'yang', 'lain', 'lain', 'nya', 'besar', 'banget'	negatif
saya pengguna baru dari bulan kemarin jika dapat penjual yang amanah dan terpercaya pasti kita happy kali belanja ada kali penjual yang tidak amanah barangnya gak seperti deskripsi dari penjual pas sampai beda	'saya', 'pengguna', 'baru', 'dari', 'bulan', 'kemarin', 'jika', 'dapat', 'penjual', 'yang', 'amanah', 'dan', 'terpercaya', 'pasti', 'kita', 'happy', 'kali', 'belanja', 'ada', 'kali', 'penjual', 'yang', 'tidak', 'amanah', 'barangnya', 'gak', 'seperti', 'deskripsi', 'dari', 'penjual', 'pas', 'sampai', 'beda'	positif

3.2.5. *Stemming*

Tahapan selanjutnya akan merubah data yang memiliki imbuhan kedalam bentuk dasar. Karena data *review* pengguna menggunakan Bahasa Indonesia, maka *stemming* akan dilakukan dengan Sastrawi. Dengan *library* ini, kata seperti "*bertahun*" akan diubah menjadi "*tahun*", "*penjual*" menjadi "*jual*" dan sebagainya seperti yang tertera dalam tabel.

Tabel 6. *Stemming*

TABLE VI. *STEMMING*

Tokenization	***Stemming***	**Label**
'sudah', 'bertahun', 'tahun', 'saya', 'belanja', 'di', 'shopee', 'sampai', 'sekarang', 'pelayanannya', 'bagus', 'sih', 'menurut', 'aq', 'baik', 'dari', 'tokonya', 'juga', 'kurir', 'nya', 'sejauh', 'ini', 'g', 'pernah', 'terjadi', 'sesuatu', 'yg', 'mengecewakan', 'sih', 'untuk', 'saya', 'sebagai', 'pelanggan'	sudah tahun tahun saya belanja di shopee sampai sekarang layan bagus sih turut aq baik dari toko juga kurir nya jauh ini g pernah jadi sesuatu yg kecewa sih untuk saya bagai langgan	positif
'akhir', 'ini', 'shopee', 'kenapa', 'ya', 'kesal', 'banget', 'sama', 'pengirimannya', 'kurir', 'ga', 'ada', 'memberi', 'kabar', 'mau', 'mengantar', 'tapi', 'kok', 'gagal', 'terus', 'dan', 'akhirnya', 'dibatalkan', 'dan', 'dikembalikan', 'kepenjual'	akhir ini shopee kenapa ya kesal banget sama kirim kurir ga ada beri kabar mau antar tapi kok gagal terus dan akhir batal dan kembali jual	negatif
'update', 'terbaru', 'malah', 'down', 'server', 'error', 'jadi', 'bagaimana', 'kita', 'mau', 'belanja', 'mau', 'masuk', 'saja', 'susah', 'banget'	update baru malah down server error jadi bagaimana kita mau belanja mau masuk saja susah banget	netral
'terlalu', 'banyak', 'biaya', 'yang', 'lain', 'lain', 'nya', 'di', 'tambah', 'lagi', ongkos, 'kirim', 'nya', 'mahal', 'banget', 'sekarang', 'harga', 'barang', 'ga', 'seberapa', 'ongkir', 'sama', 'biaya', 'yang', 'lain', 'lain', 'nya', 'besar', 'banget'	terlalu banyak biaya yang lain lain nya di tambah lagi ongkos kirim nya mahal banget sekarang harga barang ga berapa ongkir sama biaya yang lain lain nya besar banget	negatif
'saya', 'pengguna', 'baru', 'dari', 'bulan', 'kemarin', 'jika', 'dapat', 'penjual', 'yang', 'amanah', 'dan', 'terpercaya', 'pasti', 'kita', 'happy', 'kali', 'belanja', 'ada', 'kali', 'penjual', 'yang', 'tidak', 'amanah', 'barangnya', 'gak', 'seperti', 'deskripsi', 'dari', 'penjual', 'pas', 'sampai', 'beda'	saya guna baru dari bulan kemarin jika dapat jual yang amanah dan percaya pasti kita happy kali belanja ada kali jual yang tidak amanah barang gak seperti deskripsi dari jual pas sampai beda	positif

3.2.6. *Stop Words*

Setelah itu, data yang tidak memiliki arti akan dihapus seperti kata penghubung, kata ganti, dan kata umum lainnya agar memperjelas makna teks yang diproses. Contoh kata yang dihapus yaitu *"yang", "tidak", "sekali", "kenapa"*, *"lebih"* dan sebagainya. Selanjutnya, melakukan kustom untuk menghapus daftar kata yang tidak relevan seperti kata sentiment negatif yang masuk ke dalam kategori positif maupun netral atau

ISSN: 2715-6427

sebaliknya. Seperti contoh, kata "*lambat"* dan "*gagal"* akan dihapus dalam data yang berlabel positif dan netral. Data yang telah dilakukan proses *stopwords* tertera pada tabel di bawah ini.

Tabel 7. *Stop Words*

TABLE VII. *STOP WORDS*

Stop Words	**Label**
belanja shopee layan bagus toko kurir langgan	positif
shopee kesal kirim kurir kabar gagal batal jual	negatif
update server belanja masuk	netral
biaya ongkir mahal harga barang ongkir biaya	negatif
kemarin jual amanah percaya happy kali belanja kali jual amanah barang deskripsi jual beda	positif

3.3. *TF-IDF (Term Frequency-Inverse Document Frequency)*

Setelah data dilakukan tahap *preprocessing*, data lalu dibagi menjadi *data train* dan *data test* sebesar 8:2. Lalu diberikan bobot menggunakan *library python* dari *scikit-learn* yaitu *TfidfVectorizer* dengan menentukan maksimal fitur kata (*max_features)* sebesar 10000. Lalu menggunakan label encoder untuk mengubah label menjadi *one-hot encoding*. Hasil dari pembobotan *TF-IDF* dan *One-hot encoding* ditampilkan pada gambar dibawah ini.

```
Dokumen ke-84:
belanja shopee layan bagus toko kurir langgan

TF-IDF Scores:
          TF-IDF
langgan  0.498169
layan    0.424687
bagus    0.417985
toko     0.384939
kurir    0.382735
belanja  0.271701
shopee   0.168140
```

Gambar 4. Skor pembobotan *TF-IDF*

```
Label train one-hot encoded:
 [[0. 0. 1.]
 [1. 0. 0.]
 [1. 0. 0.]
 ...
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]]

Label test one-hot encoded:
 [[0. 0. 1.]
 [0. 0. 1.]
 [1. 0. 0.]
 ...
 [1. 0. 0.]
 [0. 0. 1.]
 [1. 0. 0.]]
```

Gambar 5. *One-hot encoding* pada label

3.4. *LSTM (Long Short-Term Memory)*

Sebelum menggabungkan *TF-IDF (Term Frequency-Inverse Document Frequency)* dan *LSTM (Long Short-Term Memory)*, data terlebih dahulu dilakukan tokenisasi menggunakan *TensorFlow* dari *Keras* untuk mengonversi ke dalam *sequence* atau urutan data. Panjang maksimal *sequence* yang digunakan sebesar 100.

Selanjutnya menambahkan layer *LSTM (Long Short-Term Memory)* dan menggabungkan *TF-IDF (Term Frequency-Inverse Document Frequency)* dengan *LSTM (Long Short-Term Memory)* menggunakan *concatenate* dari *Keras. Library* ini menggabungkan *output* dari *TF-IDF* dengan *LSTM*. Setelah itu menggunakan *dense* sebanyak jumlah label yaitu 3 dengan *activation* yang digunakan *softmax* untuk *multiclass*. Setelah itu melakukan *compile* model dengan *optimizer* yang digunakan yaitu *Adam*, menggunakan

Loss categorical_crossentropy untuk *multiclass*. Dan terakhir melatih model dengan *epochs* sebesar 10, dan *batch size* 64. Berikut *model summary* dan hasil pelatihan model.

```
Model: "model_3"
_________________________________________________________________________________________________
 Layer (type)                Output Shape              Param #   Connected to
=================================================================================================
 sequence_input (InputLayer  [(None, 100)]             0         []
 )

 embedding_3 (Embedding)     (None, 100, 100)          1000000   ['sequence_input[0][0]']

 lstm_3 (LSTM)               (None, 64)                42240     ['embedding_3[0][0]']

 tfidf_input (InputLayer)    [(None, 1296)]            0         []

 concatenate_3 (Concatenate  (None, 1360)              0         ['lstm_3[0][0]',
 )                                                                 'tfidf_input[0][0]']

 dense_3 (Dense)             (None, 3)                 4083      ['concatenate_3[0][0]']

=================================================================================================
Total params: 1046323 (3.99 MB)
Trainable params: 1046323 (3.99 MB)
Non-trainable params: 0 (0.00 Byte)
```

Gambar 6. *Model Summary*

```
_________________________________________________________________________________
Epoch 1/10
45/45 [==============================] - 22s 353ms/step - loss: 0.9570 - accuracy: 0.5684 - val_loss: 0.8983 - val_accuracy: 0.5778
Epoch 2/10
45/45 [==============================] - 12s 265ms/step - loss: 0.8081 - accuracy: 0.6806 - val_loss: 0.6317 - val_accuracy: 0.7854
Epoch 3/10
45/45 [==============================] - 13s 282ms/step - loss: 0.5870 - accuracy: 0.8015 - val_loss: 0.5288 - val_accuracy: 0.7854
Epoch 4/10
45/45 [==============================] - 14s 300ms/step - loss: 0.3936 - accuracy: 0.8352 - val_loss: 0.4578 - val_accuracy: 0.8050
Epoch 5/10
45/45 [==============================] - 14s 316ms/step - loss: 0.3034 - accuracy: 0.8615 - val_loss: 0.4371 - val_accuracy: 0.8079
Epoch 6/10
45/45 [==============================] - 13s 296ms/step - loss: 0.2458 - accuracy: 0.9092 - val_loss: 0.4203 - val_accuracy: 0.8205
Epoch 7/10
45/45 [==============================] - 12s 267ms/step - loss: 0.2068 - accuracy: 0.9302 - val_loss: 0.4394 - val_accuracy: 0.8261
Epoch 8/10
45/45 [==============================] - 12s 269ms/step - loss: 0.2000 - accuracy: 0.9309 - val_loss: 0.4795 - val_accuracy: 0.8233
Epoch 9/10
45/45 [==============================] - 7s 151ms/step - loss: 0.1544 - accuracy: 0.9478 - val_loss: 0.5059 - val_accuracy: 0.8247
Epoch 10/10
45/45 [==============================] - 8s 183ms/step - loss: 0.1385 - accuracy: 0.9572 - val_loss: 0.5385 - val_accuracy: 0.8317
```

Gambar 7. Pelatihan Model

3.5. Evaluasi Hasil

Setelah data dilatih, kemudian ditampilkan hasil akurasinya menggunakan *sklearn.metrics*. Berikut hasil metriks akurasi dari pelatihan ini.

```
23/23 [==============================] - 0s 19ms/step - loss: 0.5385 - accuracy: 0.8317
Test Accuracy: 0.832
23/23 [==============================] - 2s 72ms/step
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.92      0.90       412
           1       0.49      0.51      0.50        88
           2       0.89      0.80      0.84       213

    accuracy                           0.83       713
   macro avg       0.75      0.74      0.75       713
weighted avg       0.83      0.83      0.83       713
```

Gambar 8. Metriks Akurasi

Terlihat bahwa dalam *epoch* (pelatihan) terakhir model mencatat nilai *loss* (kerugian) sebesar 0.1385 dengan akurasi sebesar 83%. Presisi model dalam mengidentifikasi kelas negatif sebesar 88%, kelas netral 49%, dan kelas positif 89%. Nilai *recall* model yang berhasil mengidentifikasi kelas sebenarnya sebesar 92% untuk kelas negatif, 51% untuk kelas netral, dan 80% untuk kelas positif. Dalam pelatihan ini, model mengambil sebanyak 412 kelas negatif, 88 kelas netral, dan 213 kelas positif.

ISSN: 2715-6427

Selanjutnya, menggunakan *wordcloud* untuk mengetahui frekuensi kata yang sering dikatakan oleh pengguna. Berikut adalah hasil *wordcloud* pengguna yang terdiri dari sentimen positif, sentimen negatif, dan sentimen netral.

Gambar 9. *Word Cloud* Label Positif, Negatif, dan Netral

Dalam *word cloud* terlihat bahwa kata yang paling banyak muncul dalam kelas positif adalah *'shopee'* sebanyak 1160 kali, *'barang'* sebanyak 559 kali, *'kirim'* sebanyak 517 kali , *'belanja'* sebanyak 508 kali, dan kata *'aplikasi'* sebanyak 498 kali. Untuk kelas negatif adalah kata *'shopee'* sebanyak 1958 kali, *'kirim'* sebanyak 1360 kali, *'aplikasi'* sebanyak 1041 kali, *'barang'* sebanyak 894 kali, dan kata *'pakai'* sebanyak 539 kali. Untuk kelas netral ada beberapa kata seperti *'shopee'* sebanyak 364 kali, *'kirim'* sebanyak 258 kali, *'aplikasi'* sebanyak 191 kali, *'barang'* sebanyak 161 kali, *'belanja'* sebanyak 130 kali.

4. KESIMPULAN

Berdasarkan hasil uji, dapat ditemukan bahwa model memiliki akurasi sebesar 83% dengan kata-kata yang paling sering muncul seperti, 'shopee', 'barang', 'kirim', 'belanja', dan 'aplikasi' dalam kelas positif. Lalu kata 'shopee', 'kirim', 'aplikasi', 'barang', dan 'pakai' untuk kelas negatif, dan kata 'shopee', 'kirim', 'aplikasi', 'barang', dan 'belanja'untuk kelas netral. Model memiliki kemampuan yang cukup baik dalam memprediksi kelas negatif dan positif tetapi kurang efektif untuk kelas netral yang memiliki presisi dan recall lebih rendah. Hal ini dikarenakan data kelas netral yang kurang seimbang dibandingkan kelas yang lain. Sedangkan, *LSTM* memerlukan data yang besar agar tidak terjadi *overfitting*. Data netral yang sedikit juga dipengaruhi oleh ulasan pengguna pada *Google Play Store*. Hasil dari analisis dapat digunakan untuk memahami sentimen pengguna dan meningkatkan kualitas layanan aplikasi Adapun saran untuk penelitian selanjutnya dapat menggunakan metode yang bisa mengatasi data yang sedikit seperti *Naïve Bayes, Logistic Regression,* dan lainnya.

UCAPAN TERIMA KASIH

Terima kasih kepada semua pihak yang telah mendukung, mengarahkan, maupun membantu dalam kepenulisan jurnal penelitian ini. Terima kasih kepada dosen pembimbing atas bimbingan selama ini dan juga pengalaman serta wawasan yang dibagikan sehingga jurnal penelitian ini terselesaikan dengan baik. Terima kasih kepada orang tua, keluarga, dan teman yang mendukung dari awal hingga akhirnya jurnal ini dapat terselesaikan.

REFERENSI

[1] A. Ahdiat, "5 E-Commerce dengan Pengunjung Terbanyak Kuartal I 2023," Katadata. Accessed: Apr. 25, 2024. [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/05/03/5-e-commerce-dengan-pengunjung-terbanyak-kuartal-i-2023

[2] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, Jan. 2023, doi: 10.37905/jjeee.v5i1.16830.

[3] R. Wahyudi *et al.*, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *JURNAL INFORMATIKA*, vol. 8, no. 2, 2021, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji

[4] W. Kurnia, "Sentimen Analisis Aplikasi E-Commerce Berdasarkan Ulasan Pengguna Menggunakan Algoritma Stochastic Gradient Descent," *JURNAL TEKNOLOGI DAN SISTEM INFORMASI*, vol. 4, no. 1, pp. 138–143, 2023, doi: 10.33365/jtsi.v4i2.2561.

[5] E. Hasibuan and E. A. Heriyanto, "ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER," *JTS*, vol. 1, no. 3.

[6] Y. Tian, Y. Zhang, and H. Zhang, "Recent advances in stochastic gradient descent in deep learning," *Mathematics*, vol. 11, no. 3, p. 682, 2023.

[7] S. S. Bafjaish, "Comparative analysis of Naive Bayesian techniques in health-related for classification task," *Journal of Soft Computing and Data Mining*, vol. 1, no. 2, pp. 1–10, 2020.

[8] F. N. Fajri and S. Syaiful, "Klasifikasi Nama Paket Pengadaan Menggunakan Long Short-Term Memory (LSTM) Pada Data Pengadaan," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2635.

[9] Srivatsavaya Prudhviraju, "LSTM — Implementation, Advantages and Diadvantages," Medium. Accessed: Mar. 23, 2024. [Online]. Available: https://medium.com/@prudhviraju.srivatsavaya/lstm-implementation-advantages-and-diadvantages-914a96fa0acb

[10] M. I. Syafaah and L. Lestandy, "Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory)," 2022. [Online]. Available: https://atapdata.ai/dataset/192/HIMPUNAN_DATA_E

[11] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Front Energy Res*, vol. 9, p. 652801, 2021.
[12] A. Tholib and M. Kom, "Implementasi Algoritma Machine Learning Berbasis Web dengan Framework Streamlit," *Pustaka Nurja*, 2023.
[13] M. Z. Yumarlin, J. E. Bororing, S. Rahayu, and J. A. Putra, "Analisis Sentimen Pengguna Apliaksi Shopee Menggunakan Metode Naive Bayes Classifier dan K-NN," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 12, no. 3, pp. 745–753, 2023.
[14] N. Aula, M. Ula, and L. Rosnita, "ANALISIS SENTIMEN REVIEW CUSTOMER TERHADAP PERUSAHAAN EKSPEDISI JNE, J&T EXPRESS DAN POS INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) ANALYSIS OF CUSTOMER REVIEW SENTIMENT TO JNE, J&T EXPRESS AND POS INDONESIA EXPEDITION COMPANIES USING SVM METHOD," *Journal of Informatics and Computer Science*, vol. 9, no. 1, 2023.
[15] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J Big Data*, vol. 8, pp. 1–16, 2021.
[16] A. R. Lubis and M. K. M. Nasution, "Twitter Data Analysis and Text Normalization in Collecting Standard Word," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 4, no. 2, pp. 855–863, 2023.
[17] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.
[18] H. Benhar, A. Idri, and J. L. Fernández-Alemán, "Data preprocessing for heart disease classification: A systematic literature review," *Comput Methods Programs Biomed*, vol. 195, p. 105635, 2020, doi: https://doi.org/10.1016/j.cmpb.2020.105635.
[19] S. Shekhar, "Twitter sentiment analysis: An Arabic text mining approach based on COVID-19," *International Research Journal of Engineering and Technology*, 2021, [Online]. Available: www.irjet.net
[20] M. E. Rianto, M. Maulidiansyah, and A. Tholib, "Implementasi AI Chatbot Sebagai Support Assistant Website Universitas Nurul Jadid Menggunakan Algoritma Long Short-Term Memory (LSTM)," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 1, pp. 267–275, 2024.
[21] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi," *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, p. 012017, Jun. 2020, doi: 10.1088/1757-899X/874/1/012017.
[22] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLoS One*, vol. 16, no. 8, pp. e0254937-, Aug. 2021, [Online]. Available: https://doi.org/10.1371/journal.pone.0254937
[23] F. Nur Fajri, A. Tholib, and W. Yuliana, "Application of Machine Learning Algorithm for Determining Elective Courses in Informatics Study Program," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, Dec. 2022, doi: 10.28932/jutisi.v8i3.3990.
[24] M. M. Hoque, "An analytical approach to analyze the popular word search from nineteen-year news dataset using Natural language processing technique," Centria University of Applied Sciences, 2023.
[25] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, "The effect of the TF-IDF algorithm in times series in forecasting word on social media," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, pp. 976–984, Apr. 2021, doi: 10.11591/ijeecs.v22.i2.pp976-984.
[26] G. S. N. Murthy, S. R. Allu, B. Andhavarapu, M. Bagadi, and M. Belusonti, "Text based sentiment analysis using LSTM," *Int. J. Eng. Res. Tech. Res*, vol. 9, no. 05, 2020.
[27] A. Tholib, N. K. Agusmawati, and F. Khoiriyah, "Prediksi Harga Emas Menggunakan Metode Lstm Dan Gru," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, 2023.
[28] X. Luo and K. Tang, "Funny3 at SemEval-2020 Task 7: Humor Detection of Edited Headlines with LSTM and TFIDF Neural Network System," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, Stroudsburg, PA, USA: International Committee for Computational Linguistics, 2020, pp. 1013–1018. doi: 10.18653/v1/2020.semeval-1.132.
[29] U. O. R. Permatasari, W. J. Shudiq, and M. Jasri, "Prediksi Kelayakan Mahasiswa sebagai Penerima Beasiswa Bank Indonesia pada Tahap Seleksi Administrasi di Universitas Nurul Jadid Menggunakan Algoritma K Nearest Neighbor," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 1, pp. 252–260, 2024.
[30] A. Puji Astuti, S. Alam, and I. Jaelani, "Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo," *Jurnal Bangkit Indonesia*, vol. 11, no. 2, pp. 1–6, Oct. 2022, doi: 10.52771/bangkitindonesia.v11i2.196.