

PERBANDINGAN ALGORITMA *MACHINE LEARNING* UNTUK SENTIMEN ANALISIS *GAME BUS SIMULATOR INDONESIA*

Hery Oktafiandi ^{1*}, Winarnie ²

¹Program Studi Sistem Informasi Kota Palembang, Universitas Satu

²Program Studi Teknologi Rekayasa Perangkat Lunak, Politeknik Sawunggalih Aji

¹Rukan Taman Harapan Indah B3&B5, Jl. Letda Abdul Rozak, Palembang

²Jl. Wisma Aji No.8, Kutoarjo, Purworejo, Jawa Tengah

hery.oktafiandy@univ.satu.ac.id, winzahwa@gmail.com

Abstract - This study aims to perform sentiment analysis on user reviews of the Bus Simulator Indonesia (BUSSID) application, developed by Maleo from Surabaya and released in 2017. A total of 10,000 reviews were collected from the Google Play Store using web scraping techniques and labeled based on ratings; reviews with ratings above 3 were considered positive, while those with ratings of 3 or below were considered negative. The reviews were then processed through case folding, cleaning, tokenizing, stopword removal, and stemming, and their features were extracted using the TF-IDF method. The data was split into 70% for training and 30% for testing. Three machine learning algorithms were applied: Naive Bayes Classifier (NBC), Stochastic Gradient Descent (SGD), and Support Vector Machine (SVM). The results showed that SVM had the highest accuracy at 79%, followed by SGD at 77%, and NBC at 76%. Evaluation using metrics such as accuracy, precision, recall, and f1-score indicated that this analysis provides valuable insights for BUSSID developers to improve the application's quality.

Keywords - sentiment analysis, Naive Bayes Classifier, Stochastic Gradient Descent, Support Vector Machine, TF-IDF, Bus Simulator Indonesia

Abstrak - Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Bus Simulator Indonesia (BUSSID), yang dikembangkan oleh Maleo dari Surabaya dan dirilis pada tahun 2017. Data ulasan sebanyak 10.000 ulasan dikumpulkan dari Google Play Store melalui teknik web scraping dan dilabeli berdasarkan rating; ulasan dengan rating lebih dari 3 dianggap positif, sementara ulasan dengan rating 3 atau kurang dianggap negatif. Ulasan tersebut kemudian diproses melalui tahapan case folding, cleaning, tokenizing, penghapusan stopwords, dan stemming, serta diekstraksi fitur menggunakan metode TF-IDF. Data dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Tiga algoritma pembelajaran mesin diterapkan: Naive Bayes Classifier (NBC), Stochastic Gradient Descent (SGD), dan Support Vector Machine (SVM). Hasil menunjukkan bahwa SVM memiliki akurasi tertinggi sebesar 79%, diikuti oleh SGD dengan 77%, dan NBC dengan 76%. Evaluasi menggunakan metrik akurasi, presisi, recall, dan f1-score mengindikasikan bahwa analisis ini memberikan wawasan berharga bagi pengembang BUSSID untuk meningkatkan kualitas aplikasi.

Kata Kunci - analisis sentimen, Naive Bayes Classifier, Stochastic Gradient Descent, Support Vector Machine, TF-IDF

I. PENDAHULUAN

Perkembangan teknologi informasi telah memberikan dampak signifikan pada berbagai industri, termasuk industri permainan. Salah satu game yang menonjol di Indonesia adalah Bus Simulator Indonesia (BUSSID), yang dikembangkan oleh Maleo, seorang pengembang lokal dari Surabaya. Dirilis pada tahun 2017, BUSSID telah menjadi pilihan populer bagi para pecinta game simulator.

BUSSID menawarkan pengalaman mengemudi bus yang realistis di berbagai rute di Indonesia. Dengan fitur-fitur yang lengkap, pemain dapat merasakan sensasi mengemudi bus seolah-olah berada di jalan raya yang sesungguhnya. Dalam game ini, pemain memiliki kesempatan untuk menjelajahi

berbagai kota dan melintasi jalan-jalan yang berbeda, sambil menghadapi berbagai tantangan yang menarik.

Ulasan pengguna di *Google Play Store* merupakan sumber data yang kaya akan informasi mengenai kepuasan dan persepsi pengguna terhadap aplikasi tersebut. Analisis sentimen terhadap ulasan-ulasan ini menjadi penting untuk memahami sentimen umum dari pengguna dan untuk mengidentifikasi aspek-aspek yang memerlukan perbaikan. Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negatif.[1]

Untuk melakukan *analisis sentimen*, pertama-tama diperlukan data ulasan yang cukup. Dalam penelitian ini, data ulasan pengguna dari *Google Play*

Store akan dikumpulkan menggunakan teknik *scraping*. *Scraping* memungkinkan pengumpulan data dalam jumlah besar secara otomatis, yang dalam kasus ini adalah 10.000 ulasan pengguna.

Setelah data terkumpul, ulasan tersebut akan dilabeli berdasarkan *rating score*. Untuk kesederhanaan dan kejelasan analisis, hanya dua label yang akan digunakan: positif dan negatif. Ulasan dengan *rating* lebih dari 3 akan dianggap positif, sementara ulasan dengan *rating* 3 atau kurang akan dianggap negatif.

Penelitian ini akan menggunakan tiga algoritma pembelajaran mesin yang berbeda untuk melakukan analisis sentimen, yaitu *Naive Bayes Classifier* (NBC), *Stochastic Gradient Descent* (SGD), dan *Support Vector Machine* (SVM). Ketiga algoritma ini dipilih karena dikenal efektif dalam tugas-tugas klasifikasi teks.[2]

Naive Bayes Classifier (NBC) merupakan algoritma klasifikasi yang berbasis pada Teorema Bayes. Algoritma ini digunakan untuk memprediksi kelas dari suatu data dengan menghitung probabilitas kondisional dari fitur-fiturnya. NBC memiliki keunggulan dalam kemudahan penggunaan serta kecepatan dalam memproses data dalam jumlah besar. Meskipun sederhana, NBC seringkali memberikan hasil klasifikasi yang setara dengan algoritma yang lebih kompleks seperti *decision tree* dan *neural network*. Karena karakteristiknya yang handal dan efisien, NBC sering dipilih untuk berbagai aplikasi, termasuk dalam analisis sentimen.[3]

Stochastic Gradient Descent (SGD) adalah salah satu algoritma optimisasi yang digunakan dalam pembelajaran mesin untuk meminimalkan fungsi kerugian pada model. Algoritma ini bekerja dengan mendekati gradien yang benar dari fungsi kerugian dengan mempertimbangkan satu sampel data pada satu waktu, dan secara bertahap memperbaiki model berdasarkan gradien tersebut. SGD sering digunakan karena efisiensinya dalam mengatasi dataset besar dan kemampuannya untuk menangani pembaruan parameter model secara efisien.[4]

SVM (*Support Vector Machine*) dikenal sebagai salah satu metode pembelajaran terawasi terbaik untuk menganalisis data dan mengidentifikasi pola dalam klasifikasi. Keunggulan SVM terletak pada kemampuannya mencari *hyperplane* atau pemisah yang dapat memaksimalkan jarak (*margin*) antara kelas data, efektivitasnya dalam mengelola ruang fitur besar, serta kemampuannya untuk generalisasi yang tinggi. SVM juga efisien dan mudah dipasang berkat pemrosesan data berorientasi tugas. Namun, SVM memiliki kekurangan dalam pemilihan fitur dan pengaturan parameter yang dapat signifikan memengaruhi hasil akurasi klasifikasi. Untuk mengatasi kekurangan tersebut, metode pemilihan fitur seperti PSO (*Particle Swarm Optimization*) dapat digunakan untuk meningkatkan akurasi klasifikasi.[5]

Data ulasan akan dibagi menjadi dua bagian: 70% untuk data pelatihan (*training data*) dan 30% untuk data pengujian (*test data*). Pembagian ini memungkinkan evaluasi yang tepat terhadap kinerja model dalam memprediksi sentimen ulasan yang tidak terlihat sebelumnya.[6]

Dengan pendekatan ini, diharapkan dapat memberikan wawasan mendalam mengenai persepsi pengguna terhadap game *BUSSID*. Hasil *analisis sentimen* ini dapat digunakan oleh pengembang untuk memahami kekuatan dan kelemahan game mereka, serta untuk membuat perbaikan dan inovasi yang lebih baik di masa mendatang.

II. METODE PENELITIAN

Beberapa langkah Metodologi dari penelitian ini adalah:

A. Pengumpulan data

Data yang digunakan berasal dari ulasan Google Play Store menggunakan teknik *web scraping*. Data ulasan akan dilabeli berdasarkan *rating score* dari Google Play Store, di mana ulasan akan diberi label positif jika memiliki *rating* lebih dari 3 dan label negatif jika *rating* sama dengan atau kurang dari 3.

B. Pra-procesing

Tahapan proses pra-procesing adalah *case folding*, *cleaning*, *tokenisasi*, *penghapusan stopwords*, dan *stemming*. Setelah itu, akan diekstraksi dari teks menggunakan.[7]

C. Metode TF-IDF.

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode yang digunakan untuk merepresentasikan teks dalam bentuk numerik, sering digunakan dalam analisis teks, pemrosesan bahasa alami (NLP), dan pencarian informasi. TF-IDF membantu menentukan kepentingan suatu kata dalam sebuah dokumen relatif terhadap kumpulan dokumen lain (*corpus*).[8]

Proses TF-IDF[9]

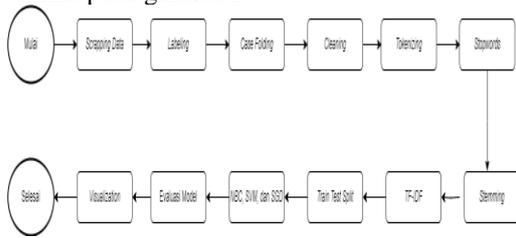
1. Preprocessing teks: terdapat proses tokenisasi (memecah teks menjadi kata-kata) dan pembersihan data (menghapus *stop words*, tanda baca, dll.).
2. Menghitung TF: Untuk setiap kata dalam dokumen, hitung frekuensinya relatif terhadap panjang dokumen.
3. Menghitung IDF: Tentukan IDF untuk setiap kata di *corpus* berdasarkan berapa banyak dokumen yang mengandung kata tersebut.
4. Menghitung TF-IDF: Kalikan nilai TF dengan IDF untuk setiap kata dalam dokumen.
5. Membentuk representasi vektor: Setiap dokumen diwakili sebagai vektor numerik yang elemen-elemennya adalah nilai TF-IDF dari kata-kata dalam *corpus*.

D. Spiting Data

Data kemudian akan dibagi menjadi dua bagian: data pelatihan (70%) dan data pengujian (30%). Algoritma *Naive Bayes Classifier* (NBC), *Stochastic Gradient Descent* (SGD), dan *Support Vector Machine* (SVM) akan diterapkan pada data pelatihan untuk melakukan analisis sentimen. [10]

E. Evaluasi model

Evaluasi model akan dilakukan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.[11] Terakhir, hasil analisis sentimen akan divisualisasikan menggunakan confusion matrix untuk memperjelas temuan dari penelitian ini. Urutan proses penelitian ini dapat dilihat pada gambar 1.



Gambar 1 Metodologi Penelitian

III. HASIL DAN PEMBAHASAN

A. Scrraping Data

Pengambilan data ulasan dari Google Play Store menggunakan google-play-scraper merupakan salah satu tahapan awal dalam proses analisis sentimen.[12] Teknik scraping ini memungkinkan pengumpulan data ulasan secara otomatis dari platform tersebut. Dalam kasus ini, data ulasan game BUSSID diambil dalam bahasa Indonesia dengan negara asal Indonesia. Penentuan urutan data berdasarkan tanggal terbaru memastikan data yang diambil lebih relevan dengan waktu penelitian. Dengan pengaturan parameter yang tepat, dihasilkan data ulasan sebanyak 10.000 yang akan menjadi basis analisis sentimen selanjutnya.

Proses scraping data ini memperluas cakupan data yang dapat diolah secara lebih luas, memungkinkan analisis yang lebih mendalam terhadap persepsi pengguna terhadap game BUSSID. Hasil dari scraping data dapat dilihat pada gambar 2.

reviewId	username	userImage	content	score	timestamp	reviewCreatedTime	at	replyContent	replyLike	replyDislike
0	be74623-3ac2-4953-8256-15420244a5d	Pengguna Google	https://play.google.com/store/apps/details?id=com.bussid	1	0	None	2024-08-10 02:58:58	None	None	None
1	4a2a01e-7164-4231-8a17-31614a31695	Pengguna Google	https://play.google.com/store/apps/details?id=com.bussid	5	0	None	2024-08-10 02:51:51	None	None	None
2	414a01e-7164-4231-8a17-31614a31695	Pengguna Google	https://play.google.com/store/apps/details?id=com.bussid	3	0	None	2024-08-10 02:43:03	None	None	None
3	4a2a01e-7164-4231-8a17-31614a31695	Pengguna Google	https://play.google.com/store/apps/details?id=com.bussid	5	0	3.71	2024-08-10 02:28:15	None	None	3.71
4	000c710-3c45-4e18-8178-304017946e	Pengguna Google	https://play.google.com/store/apps/details?id=com.bussid	4	0	4.2	2024-08-10 02:12:26	None	None	4.2

Gambar 2 Scrraping Data

Kemudian untuk proses selanjutnya diambil kolom yang akan dibutuhkan dapat dilihat pada gambar 3.

username	score	at	content
9999	Aka Yama	5 2024-04-21 04:07:38	Semoga maleo update dengan mabar menggunakan mod...
9998	Puspo Selya aji	5 2024-04-21 04:08:45	gemnya bagus busnya juga banyak
9997	Kakak Tua	1 2024-04-21 04:14:08	Gw download map Jawa Tengah g bisa ke download...
9996	Silet Silet	5 2024-04-21 04:14:31	Game bagus banget sampai aku harus ngomong fhd...
9995	S. GAMING69	5 2024-04-21 04:28:24	Bagus banget game nya mantap. Tapi untuk tim m...

Gambar 3 Data Yang Dibutuhkan

A. Labeling dan Pre-Procesing

1. Labelling

Labeling adalah proses penentuan label atau kategori untuk setiap data dalam sebuah dataset.[13] Dalam penelitian ini, proses labeling dilakukan berdasarkan skor atau rating yang diberikan oleh pengguna pada ulasan game BUSSID di Google Play Store. Ketentuan untuk labeling adalah sebagai berikut: ulasan dengan skor lebih dari 3 akan dikategorikan sebagai sentimen positif, sedangkan ulasan dengan skor 3 atau kurang akan dikategorikan sebagai sentimen negatif. Hasil dari labelling dapat dilihat pada gambar 4.

	content	sentimen
0	Semoga maleo update dengan mabar menggunakan mod...	Positif
1	gemnya bagus busnya juga banyak	Positif
2	Gw download map Jawa Tengah g bisa ke download...	Negatif
3	Game bagus banget sampai aku harus ngomong fhd...	Positif
4	Bagus banget game nya mantap. Tapi untuk tim m...	Positif

Gambar 4 Labeling

2. Case Folding

Case folding adalah salah satu tahap dalam prapemrosesan teks yang umum dilakukan dalam analisis teks.[14] Tujuan dari case folding adalah untuk mengubah semua huruf dalam teks menjadi huruf kecil atau besar, sehingga meminimalkan variasi dan memperlancar proses analisis. Dengan melakukan case folding, teks yang semula memiliki variasi huruf besar dan kecil akan disamakan, sehingga kata-kata yang sama namun ditulis dengan huruf besar atau kecil akan dianggap sama dalam analisis. Gambar 5 merupakan hasil case folding pada dataset.

	content	sentimen
0	semoga maleo update dengan mabar menggunakan mod...	Positif
1	gemnya bagus busnya juga banyak	Positif
2	gw download map jawa tengah g bisa ke download...	Negatif
3	game bagus banget sampai aku harus ngomong fhd...	Positif
4	bagus banget game nya mantap. tapi untuk tim m...	Positif

Gambar 5 Case Folding

3. Cleaning

Cleaning atau pembersihan data adalah tahap di mana teks mentah dari sumbernya dibersihkan dari karakter yang tidak relevan atau tidak diinginkan. Ini termasuk tanda baca, karakter khusus, URL, tag HTML (jika ada), dan karakter non-alfanumerik lainnya. Tujuannya adalah untuk mempersiapkan teks untuk tahap selanjutnya dalam analisis, dengan menghilangkan gangguan yang tidak perlu. Hasil dari cleaning dapat dilihat pada gambar 6.

	content	sentimen
0	semoga maleo update dengan mabar menggunakan mod	Positif
1	gemnya bagus busnya juga banyak	Positif
2	gw download map jawa tengah g bisa ke download...	Negatif
3	game bagus banget sampai aku harus ngomong	Positif
4	bagus banget game nya mantap tapi untuk tim ma...	Positif

Gambar 6 Cleaning

4. Tokenizing

Tokenizing adalah proses membagi teks menjadi unit-unit yang lebih kecil, yang disebut token. Token bisa berupa kata, frasa, atau kata-kata parsial (n-grams). Biasanya, tokenisasi dilakukan dengan memisahkan teks menjadi kata-kata berdasarkan spasi atau tanda baca. Gambar 3 merupakan hasil dari tokenizing.

	content	sentimen
0	[semoga, maleo, update, dengan, mabar, menguna...	Positif
1	[gemnya, bagus, busnya, juga, banyak]	Positif
2	[gw, download, map, jawa, tengah, g, bisa, ke,...	Negatif
3	[game, bagus, banget, sampai, aku, harus, ngom...	Positif
4	[bagus, banget, game, nya, mantap, tapi, untuk...	Positif

Gambar 7 Tokenizing

5. Stopwords

Stopwords adalah kata-kata umum yang sering muncul dalam teks dan tidak memberikan kontribusi signifikan terhadap makna atau analisis teks.[8] Contoh stopwords dalam bahasa Indonesia adalah "dan", "atau", "juga", "dari", dsb. Pada tahap ini, stopwords diidentifikasi dan dihapus dari teks, karena keberadaannya dapat mengganggu analisis teks atau klasifikasi. Hasil dari proses stopwords dapat dilihat pada gambar 8.

	content	sentimen
0	[semoga, maleo, update, mabar, menggunakan, mod]	Positif
1	[gemnya, bagus, busnya]	Positif
2	[gw, download, map, jawa, g, download, pdhl, j...	Negatif
3	[game, bagus, banget, ngomong,	Positif
4	[bagus, banget, game, nya, mantap, tim, maleo,...	Positif

Gambar 8 Stopwords

6. Stemming dan Penggabungan

Stemming yang bertujuan untuk mengonversi kata-kata dalam teks ke bentuk dasarnya atau bentuk kata dasar. Dengan menerapkan proses stemming, kata-kata yang memiliki akar yang sama dapat dikenali sebagai entitas yang sama, sehingga mengurangi redundansi dan kompleksitas dalam analisis teks. Dalam konteks bahasa Indonesia, penggunaan library Sastrawi dalam proses stemming menjadi pilihan yang umum. Library ini menawarkan algoritma stemming yang disesuaikan dengan aturan morfologi bahasa Indonesia, memastikan bahwa kata-kata dalam teks dianalisis dengan akurasi yang tinggi dan sesuai dengan struktur bahasa Indonesia.

Setelah dilakukan proses stemming, maka kata-kata tersebut akan disusun kembali menjadi kalimat untuk digunakan dalam proses selanjutnya. Gambar 9 menunjukkan hasil dari proses stemming.

	content	sentimen
0	moga maleo update mabar menggunakan mod	Positif
1	gemnya bagus bus	Positif
2	gw download map jawa g download pdhl jaring st...	Negatif
3	game bagus banget ngomong a	Positif
4	bagus banget game nya mantap tim maleo update ...	Positif

Gambar 9 Stemming dan Penggabungan

B. TF-IDF dan Train Tes Split

Frequency (TF) mengukur seberapa sering suatu istilah muncul dalam sebuah dokumen, dengan nilai TF yang semakin tinggi menandakan bahwa istilah tersebut semakin sering muncul dalam dokumen tersebut. Sedangkan Inverse Document Frequency (IDF) mengukur kelangkaan suatu istilah dalam sebuah kumpulan dokumen, dengan nilai IDF yang semakin tinggi menandakan bahwa istilah tersebut semakin jarang muncul dalam kumpulan dokumen tersebut. Gabungan antara TF dan IDF menghasilkan nilai TF-IDF, yang memberikan gambaran tentang pentingnya suatu istilah terhadap sebuah dokumen dalam sebuah korpus dokumen secara keseluruhan. Gambar 10 merupakan hasil TF-IDF beberapa kata.

```

      bagus  download  update  maleo  grafik
0  0.000000  0.000000  0.267284  0.291782  0.000000
1  0.316350  0.000000  0.000000  0.000000  0.000000
2  0.000000  0.444306  0.000000  0.000000  0.000000
3  0.233542  0.000000  0.000000  0.000000  0.000000
4  0.063179  0.000000  0.194978  0.212849  0.000000
...
9995 0.000000  0.000000  0.000000  0.222030  0.000000
9996 0.000000  0.000000  0.000000  0.000000  0.000000
9997 0.000000  0.000000  0.000000  0.000000  0.000000
9998 0.423779  0.000000  0.000000  0.000000  0.802344
9999 0.000000  0.000000  0.000000  0.000000  0.000000

[10000 rows x 5 columns]
    
```

Gambar 10 TF-IDF

Ulasan akan dipisahkan menjadi dua bagian untuk keperluan pelatihan dan pengujian model. Bagian pelatihan akan digunakan untuk membiasakan model dengan data dan pola yang terkandung di dalamnya, sementara bagian pengujian akan digunakan untuk menguji keakuratan model. Pembagian data dilakukan dengan rasio 70:30, di mana 70% digunakan untuk pelatihan dan 30% untuk pengujian. Setelah proses pembagian data, dari total 10.000 ulasan, sekitar 7.000 akan digunakan untuk pelatihan dan 3.000 untuk pengujian.

C. Klasifikasi, Evaluasi Model, dan Visualisasi

1. Stochastic Gradient Descent (SGD)

Hasil klasifikasi menggunakan algoritma SGD dengan parameter ditunjukkan pada gambar 11, gambar 12 merupakan hasil matrik pengukuran dari algoritma SGD. Visualisasi hasil konfusi matrik dapat dilihat pada gambar 13.

```

#Stochastic Gradient Descent
from sklearn.linear_model import SGDClassifier

# Inisialisasi dan latih model SGD
sgd_classifier = SGDClassifier(loss='hinge', penalty='l2', alpha=1e-3, random_state=42, max_iter=5, tol=None)
sgd_classifier.fit(X_train_tfidf, y_train)

SGDClassifier(alpha=0.001, max_iter=5, random_state=42, tol=None)
    
```

Gambar 11 Parameter SGD

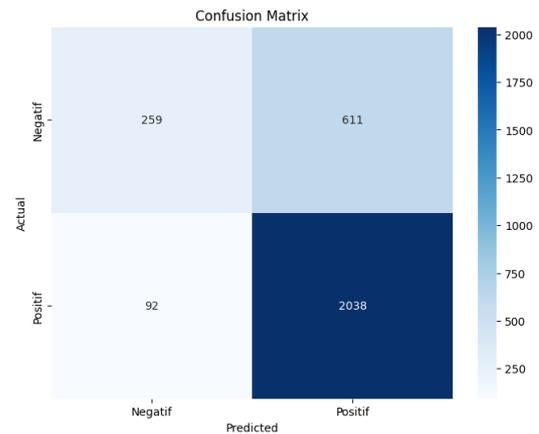
```

Accuracy: 0.7656666666666667
Confusion Matrix:
[[ 259  611]
 [  92 2038]]
Classification Report:
precision    recall  f1-score   support

 Negatif      0.74     0.30     0.42     870
 Positif      0.77     0.96     0.85    2130

 accuracy          0.77     3000
 macro avg         0.75     0.63     0.64    3000
 weighted avg      0.76     0.77     0.73    3000
    
```

Gambar 12 Hasil Matrik SGD



Gambar 13 Konfusi Matrik SGD

2. Naïve Bayes Classifier (NBC)

Hasil klasifikasi menggunakan MultinomialNB ditunjukkan pada gambar 14 yang merupakan hasil pengukuran matrik evaluasi, dan gambar 15 merupakan visualisasi dari konfusi matrik algoritma NBC.

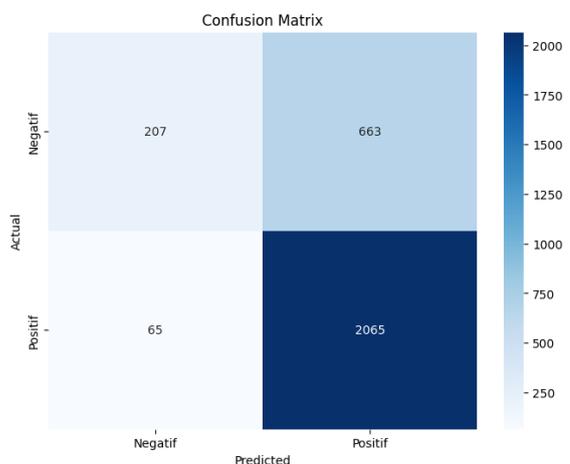
```

Accuracy: 0.7573333333333333
Confusion Matrix:
[[ 207  663]
 [  65 2065]]
Classification Report:
precision    recall  f1-score   support

 Negatif      0.76     0.24     0.36     870
 Positif      0.76     0.97     0.85    2130

 accuracy          0.76     3000
 macro avg         0.76     0.60     0.61    3000
 weighted avg      0.76     0.76     0.71    3000
    
```

Gambar 14 Hasil Matrik NBC



Gambar 15 Confosius Matrik NBC

3. Support Vector Machine (SVM)

Hasil klasifikasi menggunakan algoritma SVM menggunakan kernel linear dan random state=42 dapat dilihat pada gambar 16 yang merupakan hasil matrik evaluasi dari algoritma SVM, dan gambar 17

merupakan visualisasi dari hasil konfousius matrik algoritma SVM.

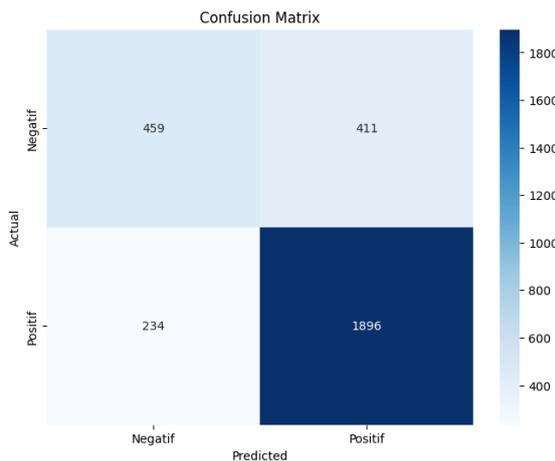
```

Accuracy: 0.785
Confusion Matrix:
[[ 459 411]
 [ 234 1896]]
Classification Report:

```

	precision	recall	f1-score	support
Negatif	0.66	0.53	0.59	870
Positif	0.82	0.89	0.85	2130
accuracy			0.79	3000
macro avg	0.74	0.71	0.72	3000
weighted avg	0.78	0.79	0.78	3000

Gambar 16 Hasil Matrik SVM



Gambar 17 Confosius Matrik SVM

Hasil klasifikasi dari dataset yang berjumlah 10000, dibagi menjadi 30% sebagai test_data dan 70% train test diperoleh nilai akurasi sebagai berikut:

Tabel 1. Hasil Akurasi Seluruh Algoritma

Algoritma	Akurasi	%
SGD	0.765	77
NBC	0.757	76
SVM	0.785	79

Dari tabel 1 dapat dilihat bahwa klasifikasi pada dataset yang sama algoritma SVM memiliki nilai akurasi yang lebih tinggi dengan nilai akurasi 79% daripada algoritma SGD yang memiliki nilai akurasi 77% dan NBC yang memiliki nilai akurasi 76%.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan model analisis sentimen untuk ulasan pengguna aplikasi Bus Simulator Indonesia (BUSSID) dengan menggunakan tiga algoritma pembelajaran mesin: Naive Bayes Classifier (NBC), Stochastic Gradient Descent (SGD), dan Support Vector Machine (SVM). Berdasarkan hasil evaluasi model, algoritma SVM menunjukkan performa terbaik dengan nilai akurasi sebesar 79%, diikuti oleh SGD dengan nilai akurasi 77%, dan NBC dengan nilai akurasi 76%. Selain itu, metrik evaluasi seperti presisi, recall, dan f1-score menunjukkan bahwa SVM memiliki kemampuan klasifikasi yang lebih unggul dibandingkan dua algoritma lainnya..

Untuk penelitian selanjutnya, disarankan untuk mempertimbangkan penggunaan teknik pemilihan fitur yang lebih canggih seperti Particle Swarm Optimization (PSO) atau Genetic Algorithm (GA) untuk lebih meningkatkan akurasi dan efisiensi model. Selain itu, penelitian lebih lanjut juga dapat mengeksplorasi penggunaan teknik deep learning seperti Recurrent Neural Networks (RNN) atau Transformer-based models untuk analisis sentimen guna mendapatkan hasil yang lebih akurat dan komprehensif. Penambahan data ulasan dari platform lain serta analisis aspek-aspek spesifik dalam ulasan juga dapat memberikan wawasan yang lebih mendalam dan bermanfaat.

DAFTAR PUSTAKA

- [1] M. H. Asnawi, I. Firmansyah, R. Novian, and R. S. Pontoh, "Comparison of Naïve Bayes, K-NN, and SVM Algorithms in Social Media Sentiment Classification," *Semin. Nas. Stat.* X, vol. 10, no. 1, 2021.
- [2] O. Irnawati and K. Solecha, "Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Berbasis Particle Swarm Optimization Pada Analisis Sentimen Ulasan Aplikasi Flip," *J. Inf. Eng. Educ. Technol.*, vol. 7, no. 1, pp. 10–15, 2023, doi: 10.26740/jieet.v7n1.p10-15.
- [3] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.
- [4] G. Ginabila and A. Fauzi, "Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine," *J. Ilm. Ilk. - Ilmu Komput. Inform.*, vol.

- 6, no. 2, pp. 111–122, 2023, doi: 10.47324/ilkominfo.v6i2.180.
- [5] A. Puji Astuti, S. Alam, and I. Jaelani, “Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo,” *J. Bangkit Indones.*, vol. 11, no. 2, pp. 1–6, 2022, doi: 10.52771/bangkitindonesia.v11i2.196.
- [6] A. J. N. Kisma, C. R. A. Widiawati, and S. Suliswaningsih, “Analysis of applications in Playstore based on Rating and Type using Naive Bayes and Logistic Regression,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 2, pp. 174–184, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [7] A. P. Prastyo and A. Y. P. Yusuf, “Sentimen Ulasan Pengguna LinkedIn Pada Aplikasi Google Playstore Dengan Metode Decision Tree,” *Jurnal Teknologi Informasi*, vol. 8, no. 1, pp. 74–82, 2024. doi: 10.36294/jurti.v8i1.3870.
- [8] H. Oktafiandi, W. Winarnie, and ..., “Perbandingan Algoritma untuk Analisis Sentimen Terhadap Google Play Store Menggunakan Machine Learning,” *J. Ekon. dan ...*, vol. 11, no. 2, pp. 16–21, 2023, [Online]. Available: <http://e-journal.polsa.ac.id/index.php/jnet/article/view/234%0Ahttp://e-journal.polsa.ac.id/index.php/jnet/article/download/234/159>
- [9] W. Kurnia, “Sentimen Analisis Aplikasi E-Commerce Berdasarkan Ulasan Pengguna Menggunakan Algoritma Stochastic Gradient Descent,” *J. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 138–143, 2023.
- [10] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm,” *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022.
- [11] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, “Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc,” *Fakt. Exacta*, vol. 14, no. 2, p. 64, 2021, doi: 10.30998/faktorexacta.v14i2.9697.
- [12] H. Junianto, P. Arsi, B. A. Kusuma, and D. I. S. Saputra, “Evaluasi Aplikasi Raileo Melalui Analisis Sentimen Ulasan Playstore Dengan Metode Naive Bayes,” *SINTECH (Science Inf. Technol. J.)*, vol. 7, no. 1, pp. 27–40, 2024, doi: 10.31598/sintechjournal.v7i1.1505.
- [13] M. F. Ravizaldi and N. Z. Saputra, “Analisis Kearifan Lokal Pada Game Bus Simulator Indonesia,” pp. 750–755, 2022.
- [14] H. Z. Muflih, A. R. Abdillah, and F. N. Hasan, “Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1613–1621, 2023, doi: 10.30865/klik.v4i3.1303.