SENTIMEN ANALISIS TERHADAP APLIKASI TIKTOK MENGGUNAKAN SUPPORT VECTOR CLASSIFICATION

Bertnaldy Christo Sidupa [1], Christine Dewi [2]
[1,2] Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
672020176@student.uksw.edu

ABSTRAK

Pemahaman terhadap persepsi pengguna aplikasi merupakan suatu aspek penting dalam pengembangan aplikasi. Sehingga menghasilkan informasi yang relefan dalam pengembangan aplikasi guna meningkatkan berbagai aspek kualitas pelayanan. Pendekatan yang digunakan dalam hal ini yakni menggunakan analisis sentimen. Perbandingan tiga metode dalam klasifikasi yang populer yakni *Random Forest, Support Vector Machine (SVM), dan Naive Bayes* merupakan topik utama dalam penelitian ini. Ulasan pengguna aplikasi *TikTok* di *Google play store* digunakan menjadi data dalam penelitian ini. Yang kemudian dimodifikasi menjadi sentimen positif, netral, dan negatif. Dalam proses ini melibatkan beberapa tahapan yakni, *pre-processing* data, pembobotan data menggunakan teknik *TF-IDF*, hingga penerapan metode klasifikasi. Selanjutnya dilakukaan evaluasi untuk mengukur kinerja setiap metode menggunakan metrik *accuracy, precission, recall,* dan *F1-score*. Dari hasil penelitian yang dilakukan teknik atau metode klasifikasi *SVM* mendapatkan hasil akurasi terbaik sebesar 85%, kemudian *Random Forest* 84%, dan *Naive Bayes* 83%. Selain mendapatkan hasil akurasi terbaik disisi lain *SVM* juga berhasil menunjukan stabilitas yang lebih baik dibandingkan metode yang lain. Penelitian ini menyimpulkan bahwa analisis sentimen terhadap aplikasi Tiktok menggunakan metode *SVM* lebih efektif. Hasil ini dapat memberikan *value*, baik dalam peningkatan aplikasi Tiktok serta peneltian selanjutnya dalam memilih metode yang terbaik dalam menganalisis data sentimen berbasis teks.

Keyword : *Analisis Sentimen, Random Forest, Support Vector Machine, Naive Bayes, Klasifikasi*

1. PENDAHULUAN

Media sosial saat ini telah menjadi salah satu sarana yang paling banyak digunakan oleh orang-orang dalam membangun interaksi dalam jaringan antara satu sama lain baik itu dalam hal saling bertukar informasi, gagasan, pandangan, serta pengalaman atau aktivitas sehari-hari. Media sosial menurut [1] merupakan media yang menjadi penghubung secara online kepada pengguna untuk menciptakan sebuah ruang komunikasi untuk berbagi , berpartisipasi dan lain-lain secara virtual. Salah satu *platform* media sosial yang saat ini menjadi *trend* ialah Tiktok. Selama masa pandemi *Covid-19* Tiktok mengalami kenaikan trend penggunaan yang dimana hal ini mempengaruhi seluruh dunia [2].

Sebagai platform media sosial yang sangat populer, TikTok juga menghadapi berbagai masalah dan tantangan yang memerlukan analisis sentimen untuk memahami respons pengguna. Menurut [3] Tiktok sempat mendapatkan sanksi dari pemerintah berupa pemblokiran karena dianggap melakukan beberapa pelanggaran sebelum kepopulerannya di Indonesia. Oleh karena itu, analisis sentimen dari *review* dan interaksi pengguna di TikTok dapat membantu dalam mendeteksi tren negatif, mengevaluasi efektivitas kebijakan moderasi konten, serta meningkatkan upaya perlindungan privasi dan keamanan pengguna. Dalam konteks aplikasi seperti TikTok, pemahaman terhadap sentimen pengguna sangat penting bagi pengembang dan pemilik aplikasi. Dengan demikian, keputusan bisnis dalam perusahaan dapat diambil dengan lebih cermat, cerdas dan tepat menggunakan hasil analasis sentimen *review* aplikasi medsos [4].

Proses mengklasifikasikan teks menjadi suatu kategori sentimen positif, netral, dan negatif dalam analisis sentimen, dengan metode klasifikasi merupakan cara yang tepat. Proses menyatakan suatu objek data menjadi kelas atau kategori adalah definisi dari sebuah klasifikasi [5]. Tiga metode klasifikasi yang umum digunakan adalah *Random Forest*, *Support Vector Machine* (*SVM*), dan *Naive Bayes*. Setiap metode klasifikasi memiliki prinsip dan pendekatan yang berbeda dalam memproses dan mengklasifikasikan data teks. Misalnya, *Random Forest* sendiri merupakan sebuah kombinasi teknik pohon keputusan yang kemudian digabung menjadi suatu model [6], selanjutnya untuk memisahkan observasi yang mempunyai target yang berbeda diperlukan sebuah fungsi pemisah *(hyperline)* yang optimal yakni menggunakan teknik *SVM* [7], dan metode Naive Bayes yaitu pengklasifikasian sederhana yang memiliki asumsi *independent* yang tinggi dari kondisi serta kejadian masing-masing yang berakar pada *Teorema Bayes* [8]. Melalui perbandingan performa ketiga metode klasifikasi ini, kita dapat mengetahui metode mana yang paling efektif dalam menganalisis sentimen pengguna terhadap aplikasi TikTok di Google Play Store. Faktor-faktor seperti akurasi, kecepatan komputasi, dan skalabilitas akan menjadi pertimbangan penting dalam mengevaluasi performa metode klasifikasi tersebut.

Dengan latar belakang ini, penelitian yang membandingkan performa ketiga metode klasifikasi dalam menganalisis sentimen aplikasi *TikTok* di Google *Play Store*. Diharapkan dapat memberikan wawasan berharga bagi pengembang aplikasi, peneliti, dan pemangku kepentingan lainnya dalam memahami respons pengguna terhadap *platform* tersebut.

2. TINJAUAN PUSTAKA

2.1. Peneltian Terdahulu

Penelitian dengan judul "Analisis Sentimen Ulasan Pengguna TikTok pada Google *Play Store* Berbasis *TF-IDF* dan *Support Vector Machine*". Penelitian yang dilakukan oleh Sukirman, Sajiah, Nursuci Putri Husain, Anastasya Febriana Syam, Ragil Mustikosari tahun 2024 merupakan penelitian yang pemrosesan datanya menggunakan metode klasifikasi SVM untuk menganalisis dan mengukur kepuasan pengguna terhadap aplikasi Tiktok yang dimana menarik kesimpulan yakni dari hasil analisis tersebut tingkat akurasi yang didapatkan yaitu 85%, presisi 79%, recall 75%, dan F1-Score sebesar 77%. Sehingga dari hasil pengujian tersebut menunjukan bahwa model yang digunakan memiliki kinerja yang solid [9].

Penelitian berjudul "Perbandingan Algoritma untuk Analisis Sentimen Terhadap Google *Play Store* Menggunakan *Machine Learning*". Penelitian yang dilakukan oleh oleh Hery Oktafiandi, Winarnie, Sayyid M. Raziq Olajuwon. Pada penelitian ini algoritma *machine learning* yang digunakan yaitu *Naive Bayes*, *K-Nearest Neighbors* dan *Random Forest* dengan tujuan untuk membandingkan algoritma mana yang memiliki hasil akurasi yang terbaik guna menganalisis sentimen terhadap aplikasi Google *Play Store*. Pada penelitian ini menggunakan dataset yang berbeda-beda. Untuk dataset dengan jumlah 1000 dataset didapatkan hasil akurasi untuk algoritma *naive bayes* 79%, *KNN* 77%, dan algoritma *Random Forest* 75% [10].

Pada penelitian yang berjudul "Perbandingan Efektivitas *Naïve Bayes* dan *SVM* dalam Menganalisis Sentimen Kebencanaan di Youtube". Penelitian yang dibuat oleh Tarissa Aura Azzahra, Nurul Anisa Sri Winarsih, Galuh Wilujeng Saraswati, Filmada Ocky Saputra, Muhammad Syaifur Rohman, Danny Oka Ratmana, Ricardus Anggi Pramunendar, Guruh Fajar Shidik. ini merupakan perbandingan antara algoritma *Naïve Bayes* dan *SVM*. Data yang digunakan yakni komentar pada aplikasi youtube terkait bencana alam. Proses ini meliputi pengumpulan data, *preprocessing* data, pembobotan *TF-IDF*, serta penerapan metode klasifikasi. *SVM* menunjukkan peforma yang lebih baik dibandingkan *Naïve Bayes* dibalik dari kekurangan dan kelebihan masing-masing metode. Dengan hasil akurasi 92% dan presisi 89% dari prediksi negatif dan 94% dari prediksi positif. Sedangkan untuk *Naive Bayes* sendiri hanya menghasilkan akurasi 79% dan presisi 91% dari prediksi negatif dan 73% dari prediksi positif Penelitian ini memberikan wawasan penting mengenai aplikasi algoritma pembelajaran mesin dalam analisis sentimen [11].

2.2 Aplikasi TikTok

TikTok merupakan sebuah aplikasi platform media sosial yang dapat memberikan ruang bagi pengguna untuk saling membangun sebuah interaksi dalam internet. Menurut [12] Tiktok diluncurkan pada bulan September tahun 2016 yang dimana Tiktok sendiri merupakan sebuah aplikasi medsos asal Tiongkok. Banyak hal yang dapat dilakukan dan didapatkan melalui Tiktok yakni hiburan yang menarik, promosi produk seperti makanan, *fashion*, kosmetik dan yang lainnya [13].

2.3 *Web Scraping*

Web scraping adalah proses pengambilan informasi dari situs website yang ada [14]. Proses ini dilakukan secara otomatis dengan menggunakan perangkat lunak atau skrip. Tujuan dari web scraping adalah untuk mengekstrak data yang terstruktur atau tidak terstruktur dari halaman web dan menyimpannya dalam format yang dapat diakses dan dianalisis lebih lanjut, seperti CSV, Excel, atau database.

2.4 Analisis Sentimen

Proses ini mengacu pada bidang yang luas dengan tujuan untuk menganalisa pendapat ataupun sentimen terhadap suatu produk, layanan, ataupun suatu kegiatan melalui pengolahan bahasa alami, komputasi *linguistic*, dan *text mining* [15].

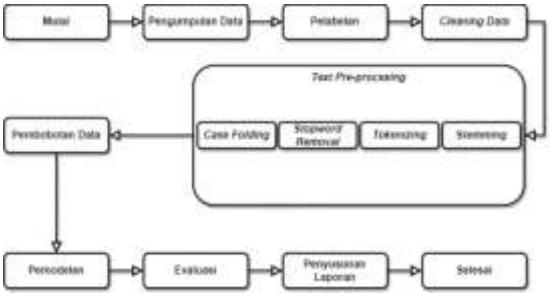
2.5 *Text mining*

Text mining adalah analisis teks yang dilakukan oleh komputer secara otomatis sehingga informasi yang didapatkan berkualitas dan terangkum dalam sebuah dokumen. Dari sekumpulan data tekstual yang besar text mining berusaha menemukan suatu pola yang menarik dengan menggunakan variasi dari data mining [16].

2.6 Klasifikasi

Klasifikasi merupakan suatu cara atau proses yang dilakukan guna memperkirakan kelas dari suatu objek yang memiliki label yang tidak diketahui dengan cara menemukan model yang menjelaskan suatu perbedaan kelas data [17]. (1) Pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan (2) Penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya adalah dua cara utama klasifikasi dalam bekerja [18].

3. METODE PENELITIAN



Gambar 1. Alur proses penelitian

3.1. Pengumpulan Data

Proses pengumpulan data pada penelitian ini menggunakan teknik atau cara *scraping* dengan *library python google play scraper.* Proses pengumpulan data web secara otomatis dan terstruktur menggunakan kode pemrograman merupakan definisi dari *Web scraping* [19]. Pada proses ini data yang diambil sejumlah 30000 data *review* dengan *sorting most relevant* pada aplikasi Tiktok di *playstore.*

3.2. Pelabelan

Pelabelan adalah suatu tahapan penting dalam analisis sentimen yang dimana proses atau tahapan ini dilakukan untuk memberikan label atau kategori yang terdapat dalam ulasan atau komentar tersebut. Pada tahapan ini proses labeling menggunakan kategori sentimen positif, netral, dan negatif.

3.3. *Cleaning Data*

Cleaning data adalah proses dalam membuat indentifikasi, analisis, mengubah, serta menghapus data yang tidak lengkap, tidak akurat, tidak relevan, atau tidak terkonsistensi dalam sebuah data set. Hal ini sangat penting sebab data yang masih kotor dan tidak terstruktur tentunya akan mempengaruhi hasil yang didapatkan.

3.4. *Text Pre-processing*

Tujuan dilakukannya *text pre-processing* ialah mempersiapkan data awal menjadi data yang benar-benar siap untuk digunakan dengan melalui beberapa tahapan didalamnya [13].

3.5. *Case Folding*

Tujuan utama dari *case folding* adalah untuk mengubah teks menjadi format yang seragam dalam hal kasus huruf, sehingga memudahkan analisis dan pemrosesan lebih lanjut tanpa memperhatikan perbedaan kasus huruf. Pada proses ini semua karakter ‘A’-’Z’ diubah menjadi karakter ‘a’-’z’.

3.6. *Stopword Removal*

Dalam analisis teks proses menghapus sebuah kata yang tidak terdapat nilai tambah didalamnya disebut dengan *stopword removal*. Kata-kata ini, seperti "dan", "atau", "yang", "dari", "ke", dan sebagainya, biasanya tidak memberikan makna khusus dalam konteks analisis teks dan sering muncul dalam teks dalam jumlah besar.

3.7. *Tokenizing*

Token merupakan suatu teks yang telah dipecah menjadi bagian-bagian kecil yang dimana proses itu disebut dengan *tokenizing*. Tujuan dari *tokenizing* adalah untuk mempermudah analisis teks dengan memungkinkan komputer untuk memahami dan memanipulasi teks secara lebih efisien.

3.8. *Stemming*

Stemming merupakan suatu pemrosesan dalam penyederhanaan kata yakni mengubah kata menjadi kebentuk dasarnya. Hal ini memiliki tujuan yakni untuk menghindari kata memiliki makna yang sama.

3.9. Pembobotan Data

Pembobotan data dalam klasifikasi adalah proses memberikan bobot atau nilai tertentu kepada fitur atau atribut dalam dataset. Tujuannya adalah untuk menyesuaikan pengaruh masing-masing fitur terhadap proses klasifikasi, sehingga fitur yang dianggap lebih penting atau relevan mendapatkan bobot yang lebih tinggi, sementara fitur yang dianggap kurang penting mendapatkan bobot yang lebih rendah. Pada penelitian ini pembobotan data dilakukan dengan cara *TF-IDF (Term Frequensy-Inverse Document Frequensy).* Menurut [14] *Term Frequensy-Inverse Document Frequency* merupakan salah satu metode pembobotan yang menggabungkan antara *term frequency (TF)* dan *inverse document frequency (IDF). Term frequency* adalah kemunculan suatu kata dalam suatu dokumen sedangkan *inverse document frequency (IDF)* adalah jumlah seluruh dokumen yang mengandung kata tertentu, maka metode *TF-IDF* ini dapatmenghitung total bobot sebuah dokumen dengan persamaan [14]:

$$tf_{df}idf_t = tf_{td} * \log\left(\frac{n}{df_t}\right) \quad (1)$$

Dimana:

$tf * idf$ = Bobot total dari kata

tf_{td} = Jumlah kemunculan kata dalam suatu dokumen

N = Total dokumen

df_t = Jumlah dari seluruh dokumen yang mengandung kata

3.10. *Random Forest*

Random Forest merupakan sebuah *ensemble learning method* yang menggabungkan beberapa *decision trees* (pohon keputusan) ke dalam sebuah *forest* dan menggunakan *output* dari semua *trees* tersebut untuk melakukan prediksi. Menurut [6]

proses prediksi yang dilakukan *Random Forest* menggunakan hasil dari tiap-tiap pohon keputusan untuk kemudian dikombinasikan sehingga menghasilkan rata-rata untuk regresi. *Random Forest* telah digunakan secara luasdalam berbagai aplikasi, termasuk klasifikasi teks, pengenalan gambar, bioinformatika, dan banyak lagi. Keunggulan dan kemampuan adaptasinya membuatnya menjadi salah satu pilihan utama dalam *machine learning*, terutama ketika memiliki *dataset* besar dan kompleks.

3.11. *Support Vector Machine*

SVM menggunakan hipotesis berupa fungsi-fungsi *linear* dalam sistem pembelajarannya yang dimana terdapat fitur yang berdimensi tinggi yang kemudian dilatih menggunakan algoritma berdasarkan pada teori optimasi [16]. Dengan tujuan utama dalam konteks klasifikasi ialah untuk mendapatkan *hyperplane* terbaik yang membagi data ke dalam kelas-kelas yang berbeda, sedemikian rupa sehingga jarak terdekat dari *hyperplane* ke titik data (disebut vektor pendukung atau *support vectors*) dari setiap kelas adalah maksimum. Dalam *SVM Linear*, setiap data pelatihan dikenal sebagai (xi, yi), di mana i = 1, 2, …, N dan xi = {xi1, xi2,…, xiq} T adalah atribut untuk data pelatihan I, yi{-1, +1} adalah Kelas Label [17].

3.12. *Naive Bayes*

Pada dasarnya, *Naive Bayes* menghitung probabilitas kelas target dari sebuah sampel data berdasarkan nilai-nilai fitur yang diamati pada sampel tersebut. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "*master*" tabel Keputusan. *Naive Bayes* dalam penggunaannya telah menjadi metode pemrosesan teks dalam klasifikasi baik itu klasifikasi dokumen, spam email, analisis sentimen dan lain sebagainya. Persamaan dari teorema Bayes adalah:

$$P(X) = \frac{P(H).P(H)}{P(X)} \quad (2)$$

Dimana:

X = Data *class* yang belum diketahui

H = Hipótesis data merupakan suatu *class spesific*

P(H|X) = Probabilitas hipótesis H berdasarkan kondisi X (Posteriori Probabilitas)

P(H) = Probabilitas hipótesis H (Prior Probabilitas)

P(X|H) = Probabilitas X berdasarkan kondisi pada hipótesis H

P(X) = Probabilitas X

3.13. Evaluasi

Pada penelitian ini tahapan evaluasi bertujuan untuk mengetahui kinerja atau performa dari model yang diusulkan. Tahapan evaluasi dalam klasifikasi analisis sentiment ini melibatkan serangkaian langkah untuk mengukur seberapa baik model klasifikasi dapat memprediksi kategori atau label yang benar dari data yang diberikan. Menurut [20] untuk mengukur peforma tolak ukur yang digunakan ialah *accuracy*, *precision*, dan *recall*. Adapun cara untuk menerapkan tolak ukur tersebut ialah dengan menerapkan metode *confusion matrix* yang terdiri dari *True positive, true negative, false positive,* dan *false negative* [20].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN} \quad (3)$$

3.14. Penyusunan Laporan

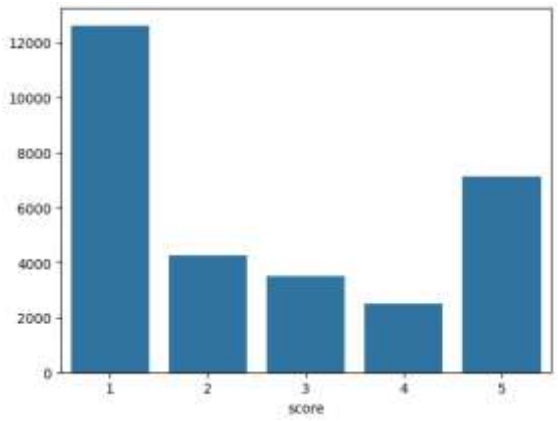
Setelah melakukan proses atau kegiatan penelitian penyusunan laporan kemudian menjadi bagian terakhir di dalamnya untuk mendokumentasikan serta mengkomunikasikan hasil penelitian yang telah dilakukan. Penyusunan laporan ini dilakukan berdasarkan data dan hasil yang ditemukan melalui proses pembelajaran materi serta pengujian.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini data yang digunakan berjumlah 30000 data yang dimana data diperoleh menggunakan teknik *web scraping* dengan alamat *scraping* yang digunakan ialah 'com.ss.android.ugc.trill'. Proses *scraping* data menggunakan bahasa pemrograman *python* pada *Google Colab*. Data yang diperoleh merupakan data komentar *user* terhadap penggunaan aplikasi Tiktok di aplikasi *Google Play Store*. Sehingga diperoleh data sebanyak 30000 data dengan 11 kolom dapat dilihat pada gambar 2 berikut.

Gambar 2. Data yang diperoleh

Kemudian data akan diidentifikasi berdasarkan *score* yang diberikan *reviewers*. Sehingga didapatkan hasil sebagai yang terlihat pada gambar 3.

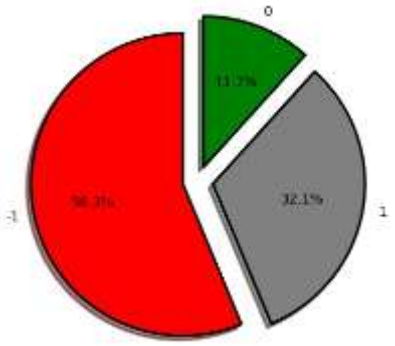


Gambar 3. *Score diagram*

Kondisi atau parameter yang digunakan ialah jika score > 3 maka sentiment akan menghasilkan nilai 1 (Puas) dan jika score == 3 maka sentiment akan menghasilkan angka 0 (Netral), dan selanjutnya jika score < 3 akan menghasilkan angka -1 (Tidak puas). Dari prosestersebut maka didapatkan hasil yaitu:

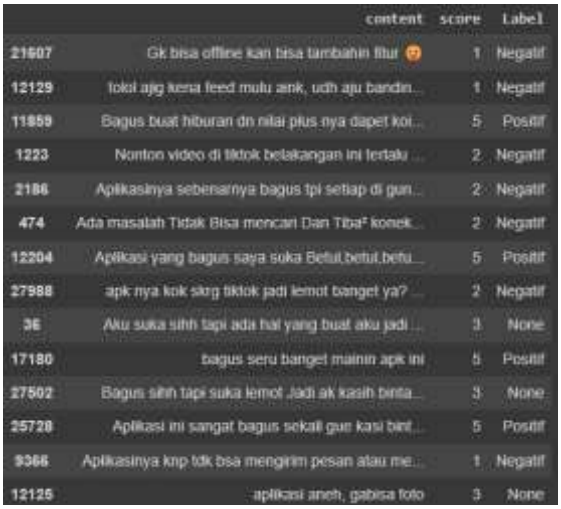
Tabel 1. Hasil identifikasi sentimen

Sentimen	Jumlah
-1	16882
1	9617
0	3501



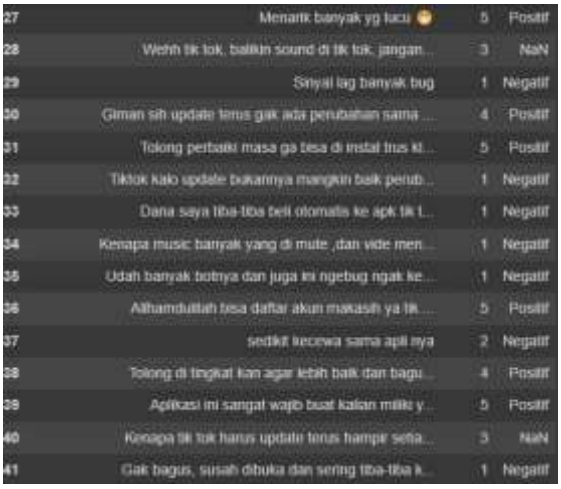
Gambar 4. *Pie chart sentiment*

Setelahnya, dilakukan proses pelabelan data dan juga proses *cleaning data* pada gambar 5 dan gambar 6. Data akan diberikan label atau kategori untuk mengidentifikasi sentimenyang didapatkan. Kemudian data yang telah dilabeling akan dilakukan pembersihan data yaitu, mengubah/menghapus data yang tidak lengkap, tidak akurat, tidak relevan, atau tidak ter konsistensi dalam sebuah dataset.



	content	score	Label
21607	Gk bisa offline kan bisa tambahin fitur	1	Negatif
12129	tokol ajig kesa feed mulu aink, udh aju bandin...	1	Negatif
11859	Bagus buat hiburan dn nilai plus nya dapet koi...	5	Positif
1223	Nonton video di tiktok belakangan ini tertalu ...	2	Negatif
2186	Aplikasinya sebenarnya bagus tpi setiap di gun...	2	Negatif
474	Ada masalah Tidak Bisa mencari Dan Tiba² koneK...	2	Negatif
12204	Aplikasi yang bagus saya suka Betul,betul,betu...	5	Positif
27988	apk nya kok skrg tiktok jadi lemot banget ya? ...	2	Negatif
36	Aku suka sihh tapi ada hal yang buat aku jadi ...	3	None
17180	bagus seru banget mainin apk ini	5	Positif
27502	Bagus sihh tapi suka lemot ,Jadi ak kasih bint...	3	None
25728	Aplikasi ini sangat bagus sekali gue kasi bint...	5	Positif
9366	Aplikasinya knp tdk bsa mengirim pesan atau me...	1	Negatif
12125	aplikasi aneh, gabisa foto	3	None

Gambar 5. Data hasil pelabelan



27	Menarik banyak yg lucu	5	Positif
28	Wehh tik tok, balikin sound di tik tok, jangan...	3	NaN
29	Sinyal lag banyak bug	1	Negatif
30	Giman sih update terus gak ada perubahan sama ...	4	Positif
31	Tolong perbaiki masa ga bisa di instal trus kl...	5	Positif
32	Tiktok kalo update bukannya mangkin baik perub...	1	Negatif
33	Dana saya tiba-tiba beli otomatis ke apk tik t...	1	Negatif
34	Kenapa music banyak yang di mute ,dan vide men...	1	Negatif
35	Udah banyak botnya dan juga ini ngebug ngak ke...	1	Negatif
36	Alhamdulilah bisa daftar akun makasih ya tik ...	5	Positif
37	sedikit kecewa sama apli nya	2	Negatif
38	Tolong di tingkat kan agar lebih baik dan bagu...	4	Positif
39	Aplikasi ini sangat wajib buat kalian milik y...	5	Positif
40	Kenapa tik tok harus update terus hampir setia...	3	NaN
41	Gak bagus, susah dibuka dan sering tiba-tiba k...	1	Negatif

Gambar 6. Hasil *cleaning data*

Kemudian proses dilanjutkan dengan tahapan *preprocessing* untuk mempersiapkan data menjadi lebih baik. Proses ini melewati beberapa tahapan yakni, *case folding* yaitu pada proses ini semua karakter ‘A’-’Z’ diubah menjadi karakter ‘a’-’z’, *stopword removal* dalam pemrosesan teks adalah proses menghapus kata-kata umum yang tidak memberikan nilai tambah dalam analisis teks. Kata-kata ini, seperti "dan", "atau", "yang", "dari", "ke", dan sebagainya, *Tokenizing* dalam pemrosesan teks adalah proses memecah teks menjadi bagian-bagian yang lebih kecil, yang disebut token, dan *stemming* dalam pemrosesan teks adalah proses menghapus awalan atau akhiran kata sehingga hanya menyisakan bentuk dasar atau kata dasar (*stem*) dari kata tersebut. Berikut data hasil pemrosesan *text pre-processing*.

Gambar 7. Data hasil pre-processing

Data kemudian dipecah menjadi 20% untuk *datatest.* Setelahnya dilakukan pembobotan data menggunakan *TF-IDF (Term Frequensy-Inverse Document Frequensy).* Data hasil pembobotan kemudian dilakukan pemodelan dengan menggunakan metode *Randomforest, Support Vector Machines*, dan juga *Naïve bayes.* Sehingga mendapatkan hasilseperti pada tabel 2, tabel 3, dan tabel 4.

Tabel 2. Hasil metode random forest

	Accuracy	**Precision**	**Recall**	**F1-Score**
AdaBoost Classifier	82%	83%	90%	86%
Bagging Classifier	82%	82%	91%	87%
RandomForest Classifie	84%	82%	95%	88%

Tabel 3. Hasil metode support vector machine

	Accuracy	**Precision**	**Recall**	**F1-Score**
LinearSVC	82%	85%	87%	86%
NuSVC	83%	81%	95%	88%
SVC	85%	84%	94%	89%

Tabel 4. Hasil metode naive bayes

	Accuracy	**Precision**	**Recall**	**F1-Score**
BernoulliNB	83.8%	86%	88%	87%
ComplementNB	83.9%	85%	89%	87%
MultinomialNB	83.8%	85%	90%	87%

Tabel 5. Hasil perbandingan metode-metode yang digunakan

	Accuracy	**Precision**	**Recall**	**F1-Score**
Random Forest (Random Forest Classifier)	84%	82%	95%	88%
Support Vector Machine (SVC)	85%	84%	94%	89%
Naive Bayes (ComplementNB)	83.9%	85%	89%	87%

Tabel 5 merupakan hasil dari ketiga metode klasifikasi yang dibandingkan dalam penelitian ini yang menunjukan bahwa metode klasifikasi *Support Vector Machines* menggunakan parameter *(SVC)* mendapatkan hasil terbaik.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang dilakukan pada permasalahan penelitian di atas didapatkan hasil bahwa tiap metode yang digunakan memiliki kelebihan masing-masing sehingga mendapatkan hasil yang bagus dan bermanfaat. Baik itu untuk metode *Random Forest, SVM, dan Naïve bayes*. Adapun untuk hasil identifikasi sentimen mendapatkan hasil positif 32.1%, negatif 56.3%, dan netral 11.7%. Untuk metode *Random Forest* dalam menganalisis masalah diatas diperoleh hasil *Accuracy* 84%, *Precision* 82%, *Recall* 95%, dan *f-1 score* 88%. Kemudian untuk metode *SVM* mendapatkan hasil *Accuracy* 85%, *Precision* 84%, *Recall* 94%, dan *F1-score* 89%. Sedangkan untuk metode *Naïve Bayes* memperoleh hasil yakni *Accuracy* 83%, *Precision* 85%, *Recall* 89%, dan *F1-score* 87%. Dengan hasil pengujian tersebut dapat disimpulkan bahwa metode yang paling baik dalam melakukan analisis data dalam studi kasus sentimen terhadap aplikasi Tiktok ialah metode klasifikasi *SVM (Support Vector Machines).*

DAFTAR PUSTAKA

[1] N. Aprilia, B. Permadi, F. A. I. A. Berampu, and S. A. Kesuma, "Media Sosial Sebagai Penunjang Komunikasi Bisnis di Era Digital," *UTILITY: Jurnal Ilmiah Pendidikan Dan Ekonomi*, vol. 7, no. 02, pp. 64–74, 2023.

[2] C. S. M. Hutajulu, S. Sherly, and H. Herman, "Peran Aplikasi Tiktok Terhadap Minat Belajar Siswa SMA," *Edukatif: Jurnal Ilmu Pendidikan*, vol. 4, no. 2, pp. 3002–3010, 2022.

[3] S. Fide, S. Suparti, and S. Sudarno, "Analisis sentimen ulasan aplikasi tiktok di google play menggunakan metode support vector machine (svm) dan asosiasi," *Jurnal Gaussian*, vol. 10, no. 3, pp. 346–358, 2021.

[4] D. Ardiansyah, A. Saepudin, R. Aryanti, and E. Fitriani, "Analisis Sentimen Review pada Aplikasi Media Sosial Tiktok Menggunakan Algoritma K-NN dan SVM Berbasis PSO," *Jurnal Informatika Kaputama (JIK)*, vol. 7, no. 2, pp. 233–241, 2023.

[5] E. T. Handayani and A. Sulistiyawati, "Analisis Setimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes," *Jurnal Teknologi Dan Sistem Informasi*, vol. 2, no. 3, pp. 32–37, 2021.

[6] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *Jurnal Informatika Upgris*, vol. 7, no. 1, 2021.

[7] H. Syah and A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm),"

Jurnal Sistem Informasi Dan Informatika (Simika), vol. 5, no. 1, pp. 59–67, 2022.

[8] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022.

[9] N. P. Husain, S. Sukirman, and S. SAJIAH, "Analisis Sentimen Ulasan Pengguna Tiktok pada Google Play Store Berbasis TF-IDF dan Support Vector Machine," *Journal of System and Computer Engineering (JSCE)*, vol. 5, no. 1, pp. 91–102, 2024.

[10] H. Oktafiandi, W. Winarnie, and S. M. R. Olajuwon, "Perbandingan Algoritma untuk Analisis Sentimen Terhadap Google Play Store Menggunakan Machine Learning," *Jurnal Ekonomi dan Teknik Informatika*, vol. 11, no. 2, pp. 16–21, 2023.

[11] T. A. Azzahra *et al.*, "Perbandingan Efektivitas Naïve Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, pp. 312–322, 2024.

[12] A. A. N. B. J. Dewanta, "Pemanfaatan Aplikasi Tik Tok Sebagai Media Pembelajaran Bahasa Indonesia," *Jurnal Pendidikan dan Pembelajaran Bahasa Indonesia*, vol. 9, no. 2, pp. 79–85, 2020.

[13] A. N. Sa'adah, A. Rosma, and D. Aulia, "Persepsi generasi Z terhadap fitur Tiktok Shop pada aplikasi Tiktok," *Transekonomika: Akuntansi, Bisnis Dan Keuangan*, vol. 2, no. 5, pp. 131–140, 2022.

[14] C. G. Indrayanto, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, pp. 1131–1139, 2023.

[15] A. Z. Amrullah, A. S. Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal Bumigora Information Technology (BITe)*, vol. 2, no. 1, pp. 40–44, 2020.

[16] D. A. C. Rachman, R. Goejantoro, and F. D. T. Amijaya, "Implementasi Text Mining Pengelompokkan Dokumen Skripsi Menggunakan Metode K-Means Clustering," *EKSPONENSIAL*, vol. 11, no. 2, pp. 167–174, 2021.

[17] S. Febriani and H. Sulistiani, "Analisis Data Hasil Diagnoga Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4. 5," *Jurnal Teknologi Dan Sistem Informasi*, vol. 2, no. 4, pp. 89–95, 2021.

[18] H. F. Putro, R. T. Vulandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN)*, vol. 8, no. 2, 2020.

[19] D. Rudini, D. G. Purnama, and A. A. Khan, "Penggunaan Teknik Web Scraping dalam Aplikasi Pengambilan Data dari Google Maps untuk Menunjang Digital Marketing," *Lentera: Multidisciplinary Studies*, vol. 2, no. 1, pp. 10–19, 2023.

[20] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 785–795, 2022.