

ANALISIS SENTIMEN *REVIEW* PENGGUNAAN *TIKTOK* MELALUI PENDEKATAN ALGORITMA *NAÏVE BAYES*

Nur Apriliani, Nana Suarna, Willy Prihartono

Teknik Informatika, Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
nuraprln182@gmail.com

ABSTRAK

Media sosial merupakan platform digital yang memungkinkan penggunanya berinteraksi dengan orang lain, berbagi konten seperti teks, gambar, video dan audio, serta terlibat dalam berbagai bentuk komunikasi online yang berbeda-beda, salah satu bentuk media sosial adalah *TikTok*. *TikTok* merupakan platform media sosial yang semakin populer yang memungkinkan pengguna berbagi video pendek dengan konten berbeda. Dengan pertumbuhan yang pesat tersebut, muncul permasalahan penting terkait pemahaman dan pengelolaan emosi pengguna terhadap konten di *TikTok*. Tujuan penelitian ini adalah melakukan analisis sentimen terhadap tinjauan penggunaan *TikTok* dengan menggunakan pendekatan algoritma *Naive Bayes*. Dalam penelitian ini, kami mengumpulkan data evaluasi dari pengguna *TikTok*, termasuk peringkat, komentar, dan opini terkait pengalaman pengguna di platform tersebut. Kami menggunakan algoritme *Naive Bayes* untuk mengklasifikasikan ulasan pengguna *TikTok* ke dalam dua kategori opini utama: positif dan negatif. Tahapan yang dilakukan meliputi pengolahan data, ekstraksi fitur dan pelatihan model *Naive Bayes* menggunakan data pelatihan. Hasil klasifikasi penggunaan metode algoritma *Naive Bayes* dan ekstraksi fitur dengan TF-IDF menghasilkan nilai akurasi sebesar 80%, precision sebesar 79%, recall sebesar 74% dan F1-Score sebesar 75%.

Kata kunci : *TikTok*, Analisis Sentimen, Algoritma *Naive Bayes*, Ulasan Pengguna.

1. PENDAHULUAN

Di era digital yang terus berkembang, platform media sosial telah menjadi bagian penting dari kehidupan sehari-hari masyarakat. Salah satu platform terpopuler saat ini adalah *TikTok*. *TikTok* adalah aplikasi berbagi video yang memungkinkan pengguna membuat dan berbagi video pendek dengan musik latar yang menarik. Dengan penggunaannya yang terus berkembang, *TikTok* telah menjadi sarana ekspresi diri, hiburan, dan interaksi sosial. *TikTok* menjadi salah satu media sosial yang marak digandrungi pengguna internet di dunia.

Permasalahan yang terjadi dalam lingkungan yang berpusat pada pengguna ini, penting bagi penyedia layanan seperti *TikTok* untuk memahami pandangan dan opini pengguna tentang platform mereka.

Penelitian sebelumnya menurut Apriani dkk dalam jurnalnya yang berjudul Analisis Sentimen Dengan *Naive Bayes* Terhadap Komentar Aplikasi Tokopedia pada tahun 2019 menyatakan bahwa algoritma *naive bayes* dapat melakukan analisis sentimen dengan benar dan melakukan klasifikasi secara otomatis setelah melalui tahapan-tahapan proses. Tahapan proses tersebut dapat berjalan dengan baik dan mengklasifikasikan data dengan parameter positif dan negatif [1].

Tujuan utama analisis sentimen adalah mengklasifikasikan teks atau ulasan ke dalam kategori sentimen yang relevan seperti positif dan negatif. Dalam hal ini, pendekatan algoritma *naive bayes* banyak digunakan dalam analisis sentimen karena sederhana, cepat dan efisien. Metode Penelitian ini

menggunakan pendekatan Algoritma *Naive Bayes* Classifier.

Hasil yang didapatkan dari algoritma *Naive Bayes* masih bekerja dengan baik dalam banyak kasus dan relatif mudah diimplementasikan. Algoritma ini telah berhasil digunakan di berbagai aplikasi, termasuk analisis sentimen ulasan produk, layanan, dan konten di media sosial. Penelitian mengenai keberhasilan penerapan analisis sentimen *TikTok* menggunakan algoritma *Naive Bayes* berpotensi memberikan wawasan tentang preferensi dan respons pengguna terhadap konten di platform.

2. TINJAUAN PUSTAKA

2.1. *TikTok*

TikTok merupakan salah satu platform media sosial yang memberikan kemungkinan bagi para penggunanya untuk dapat membuat video pendek dengan durasi hingga 3 menit yang didukung dengan fitur musik, filter, dan berbagai fitur kreatif lainnya. Awalnya *TikTok* tidak diluncurkan dengan nama *TikTok*. Pada tanggal September 2016, perusahaan *ByteDance* yang berbasis di Tiongkok meluncurkan aplikasi video pendek bernama *Douyin*. *Douyin* dapat menjangkau 100 juta pengguna dan 1 miliar penayangan video per hari hanya dalam tahun. Karena popularitas yang berkembang pesat, *ByteDance* memutuskan untuk memperluas jangkauan *Douyin* diluar Tiongkok dengan nama baru yaitu *TikTok* [2].

2.2. *Google Play*

Google Play adalah toko aplikasi resmi *Google* untuk perangkat yang menjalankan sistem operasi *Android*. *Google Play* pertama kali dirilis pada tanggal

22 Oktober 2008 dengan nama sebelumnya *Android Market* dan diubah menjadi *Google Play Store* pada bulan Maret 2012. Layanan *Google Play* sendiri dapat diakses melalui situs *websitenya*, aplikasi *Android*, dan *Google TV* [3].

2.3. Analisis Sentimen

Analisis sentimen adalah proses penggunaan pendekatan NLP (*natural Language Processing*) untuk mencari makna pandangan, pendapat, atau emosi yang diungkapkan dalam teks, ucapan, postingan (aktivitas di Internet), atau database [4]. Dengan menganalisis atau menentukan pendapat orang lain dan mengkategorikan sikap mereka terhadap suatu subjek sebagai positif, negatif, atau netral dianggap analisis sentimental [5].

2.4. Online Review

Ulasan pelanggan online adalah sejenis promosi dari *electronic word of mouth*(eWOM). Menurut Khammash (2008: 79), ulasan pelanggan online dapat dipahami sebagai suatu jenis media. Konsumen membaca ulasan dari konsumen lain tentang produk, jasa, dan metode produksi perusahaan tersebut. Menurut Li, N. dan Zhang, P. (2002), review diartikan sebagai evaluasi tentang pengalaman menyukai suatu produk, berkaitan dengan keadaan psikologis dan emosional yang dialami pengguna saat berinteraksi dengan produk tersebut [6].

2.5. Text Mining

Text Mining adalah algoritma yang mengambil informasi dari teks terstruktur. *Text mining* juga dapat membantu pengguna menemukan isi bacaannya tanpa harus membaca keseluruhan isinya. Pengelolaan sumber data pada text mining dilakukan dalam beberapa langkah. Fase ini disebut pra pemrosesan teks. Pemrosesan awal teks terdiri dari beberapa tahap: transformasi kasus, pemfilteran stopword, tokenisasi, dan stemming [7].

2.6. Naive Bayes

Metode yang digunakan untuk klasifikasi teks pada analisis sentimen adalah Naive Bayes classifier, dimana keputusan kelas pada classifier tersebut didasarkan pada probabilitas [5]. Keuntungan menggunakan Naive Bayes adalah metode ini hanya memerlukan sejumlah kecil data pelatihan (training data) untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi, sehingga lebih tepat digunakan pada situasi dunia nyata yang kompleks [8]. Karena kesederhanaannya, Naive Bayes menjadi model pembelajaran mesin generatif yang memungkinkan semua atribut berkontribusi secara setara terhadap keputusan akhir. Kesederhanaan ini setara dengan efisiensi komputasi, menjadikan Naive Bayes menarik dan cocok untuk berbagai bidang [9].

2.7. Klasifikasi

Klasifikasi adalah proses mengidentifikasi dan mengkategorikan data sebagai konsep atau kelas memerlukan penggunaan model/fungsi, seperti penyeimbangan otomatis ke kelas objek yang tidak diketahui. Modelnya sendiri bisa dalam bentuk aturan “jika-maka”, pohon keputusan, rumus matematika, atau jaringan saraf. Proses klasifikasi biasanya dibagi menjadi dua tahap: pelatihan dan pengujian. Selama fase pelatihan, beberapa data dengan kelas data yang diketahui dimasukkan ke dalam model estimasi. Kemudian pada tahap pengujian, model yang dibuat diuji dengan menggunakan data lain untuk mengetahui keakuratan model. Dengan akurasi yang cukup, model ini dapat digunakan untuk memprediksi kelas data yang tidak diketahui [10].

2.8. Scraper Data Instan

Scraper Data Instan adalah ekstensi *browser Chrome* yang dapat mengekstrak data dari halaman web dan menyimpannya sebagai file Excel atau CSV (Webrobots.io, 2022). Dengan menggunakan kecerdasan buatan, program ini dapat mengidentifikasi informasi paling relevan di halaman HTML dan mentransfernya ke file Excel atau CSV. Ini tidak memerlukan skrip khusus untuk situs web, melainkan menggunakan kecerdasan buatan untuk menganalisis struktur HTML guna memastikan informasi apa yang perlu diekstraksi. Selain itu, Pengikis Data Instan tidak memerlukan pengetahuan pemrograman, JSON, atau XML apa pun [11].

2.9. Google Collaboration

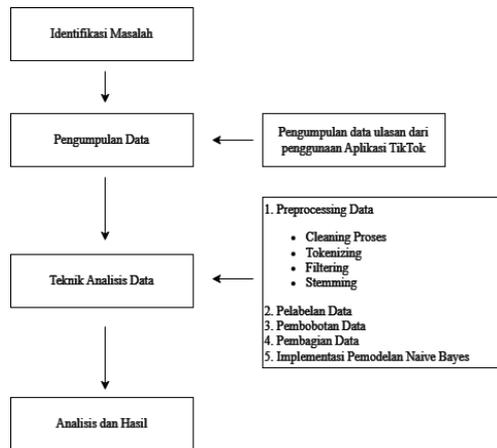
Google Colab atau *Google Colaboratory* adalah dokumen yang dapat dieksekusi yang dapat digunakan untuk menyimpan, membakar, dan berbagi program yang dibuat melalui Google Drive. Beberapa keunggulan Google Colab adalah namanya sendiri "Berkolaborasi", kemampuan untuk berkolaborasi dengan pengguna lainnya melalui berbagai pengkodean online, dan GPU gratis Google Colab yang memungkinkan pengguna melakukannya adalah kemudahan yang dapat dilakukan oleh program komputer ditulis. Spek tinggi (GPU Tesla, RAM - 12 GB, Harddisk - 300 GB) Fleksibel dan mampu menjalankan program deep learning tanpa kendala [12].

3. METODE PENELITIAN

Dalam penelitian ini, metode yang akan digunakan adalah Algoritma *Naive Bayes* (NB). *naive bayes* adalah jenis algoritma yang digunakan untuk memprediksi sentimen suatu teks atau dokumen, apakah itu positif, negatif, atau netral. Dalam konteks analisis sentimen, algoritma *naive bayes* memproses teks atau data kata-kata untuk memprediksi kategori sentimen berdasarkan probabilitas kondisional.

Pada Gambar 1.1 dibawah merupakan tahapan dalam bentuk *flowchart* yang digunakan untuk menggambarkan langkah-langkah suatu penelitian

atau proses. Diagram alir dapat membantu peneliti atau pembaca memahami urutan dan hubungan setiap langkah penelitian secara visual.



Gambar 1. Metode Penelitian

Kegiatan penelitian akan dilakukan sebanyak 4 tahapan. Dari diagram diatas dapat digambarkan tentang uraian dari masing-masing tahapan, sebagai berikut:

Tabel 1. Tahapan Penelitian

Tahap	Aktivitas	Deskripsi Aktivitas
Identifikasi Masalah	Mengidentifikasi dan mendefinisikan kebutuhan analisis sentimen penggunaan <i>TikTok</i>	Pada tahap ini dilakukan analisis terhadap kebutuhan terkait sentimen penggunaan <i>TikTok</i>
Pengumpulan data	Pengumpulan data dari ulasan pengguna <i>TikTok</i> pada <i>website Google Play</i>	Proses pengumpulan data dan input dalam sistem ini dimulai dengan data <i>scraping</i> , proses <i>scraping</i> yang dilakukan melalui <i>Data Scraper</i> , salah satu produk <i>Google Chrome</i> .
Teknik Analisis Data	Dalam menganalisis data ulasan penggunaan Aplikasi <i>TikTok</i> dalam penerapan <i>text mining</i> ini menggunakan proses tahapan <i>Knowledge Discovery in Database (KDD)</i>	<ol style="list-style-type: none"> 1. <i>Pre-Processing Data</i> <ul style="list-style-type: none"> • <i>Cleaning Proses</i> • <i>Tokenizing</i> • <i>Filtering</i> • <i>Stemming</i> 2. Pelabelan Data 3. Pembobotan 4. Pembagian Data 5. Implementasi Pemodelan dengan <i>naive bayes</i>
Analisis dan Hasil	Analisis dan Hasil	Pada tahap analisis dan hasil penelitian ini, algoritma <i>naive bayes</i> digunakan sebagai klasifikasi, suatu bentuk klasifikasi

Tahap	Aktivitas	Deskripsi Aktivitas
		probabilistik berdasarkan teorema <i>Bayes</i> (dengan asumsi kuat (<i>naive</i>) independensi statistik <i>Bayesian</i>). Aturan model yang lebih deskriptif, yaitu "model dengan karakteristik model sebelumnya".

4. HASIL DAN PEMBAHASAN

4.1. Data

Pada penelitian ini dilakukan proses pengumpulan data, dimana data yang diambil merupakan data ulasan dari pengguna aplikasi *TikTok* pada *website Google Play* yang diambil sejak tanggal 27 Agustus 2023 hingga 19 November 2023. Proses pengambilan data dilakukan dengan menggunakan teknik *scraping* melalui *Data Scraper* yang dimiliki oleh *Google Chrome*. Hasil dari proses *scraping* yang telah dilakukan disimpan dalam format Excel File yang menghasilkan jumlah ulasan atau komentar sebesar 1080 ulasan dengan 4 kolom. Hasil dari *scraping* data dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2. Hasil Scraping

No	Nama Pengguna	Tanggal	Komentar
1	Rizky Maulana	15-Nov-23	Ini tiktok kenapa sih semakin di update makin berat aplikasinya, baru masuk ke aplikasinya lama banget malah nge stuck dilogo? aplikasinya menyebabkan lag, stuck dilogo, force close, kadang suka log out sendiri? tolongdiperbaiki!!!!!!!!!!!!!! !!
2	Kadek Resta	17-Nov-23	uhh apknya seru. seru bgt malah. tapi saya sedikit mengalami bug seperti tanda "posting ulang" kadang hilang, saya agak panik karna nyaman bgt ada fitur posting ulang.
..
1080	Princess Kirana	06-Sep-23	Bgs sih, tpi kdng klu mau buka gabisa trs, trs jg lama. kalau mau buka harus hapus abis itu download lagi, gatau masalah ny kenapa jaringan bagus tapi gabisa dibuka. gatau deh cape.

4.2. Preprocessing Data

Data yang diperoleh dari *scraping* data melalui *Data Scraper* tidak selalu dalam kondisi ideal untuk diproses. Terkadang data tersebut memiliki berbagai

permasalahan yang dapat mempengaruhi hasil dari proses mining itu sendiri, seperti: *missing value*, *redundant* data, anomali atau format data yang tidak kompatibel dengan sistem. Oleh karena itu, diperlukan langkah pra-pemrosesan untuk mengatasi permasalahan tersebut. *Preprocessing* merupakan salah satu langkah untuk menghilangkan masalah yang dapat mempengaruhi hasil pengolahan data. Adapun tahapan *preprocessing* dilakukan sebagai berikut:

4.2.1. Case Folding

Pada tahap *preprocessing* terdapat bagian *case folding*, tujuan dari tahap bagian ini adalah untuk mengubah huruf kapital menjadi huruf kecil serta menghilangkan tanda baca yang ada dalam kalimat ulasan. Hasil dari proses *case folding* dapat dilihat pada tabel 3.

Tabel 3. Hasil Tahap *Case Folding*

Sebelum	Sesudah
Ini tiktok kenapa sih semakin di update makin berat aplikasinya,baru masuk ke aplikasi nya lama banget malah nge stuck dilogo?aplikasinya menyebabkan lag,stuck dilogo,force close,kadang suka log out sendiri?tolong diperbaiki!!!!!!!!!!!!!!	ini tiktok kenapa sih semakin di update makin berat aplikasinya,baru masuk ke aplikasi nya lama banget malah nge stuck dilogoaplikasinya menyebabkan lagstuck dilogoforce closekadang suka log out sendiritolong diperbaiki

4.2.2. Tokenizing

Tahap *tokenizing* merupakan suatu proses pemisahan teks kedalam dokumen menjadi potongan-potongan kata yang saling berpengaruh. Kata-kata tersebut menggunakan tanda koma (,) sebagai pemisahannya. Hasil dari proses *tokenizing* dapat dilihat pada tabel 4.

Tabel 4. Hasil Tahap *Tokenizing*

Sebelum	Sesudah
saya kira hp saya yang rusak ternyata banyak yg sama keluhannya setelah diperbarui malah setiap kali buka aplikasi tik tok ngelag harus uninstal dan download lagi baru bisa dibuka lagi aplikasinya dan itu mempengaruhi aplikasi lain jadi ngelag semua	saya, kira, hp, saya, yang, rusak, ternyata, banyak, yg, sama, keluhannya, setelah, diperbarui, malah, setiap, kali, buka, aplikasi, tik, tok, ngelag, harus, uninstal, dan, download, lagi, baru, bisa, dibuka, lagi, aplikasinya, dan, itu, mempengaruhi, aplikasi, lain, jadi, ngelag, semua

4.2.3. Filtering

Filtering kata dalam *preprocessing* data merujuk pada proses mengidentifikasi, memilih, atau menghapus kata-kata tertentu dari suatu teks atau dokumen. Tujuan *filtering* kata adalah untuk membersihkan data teks dari elemen yang dianggap tidak relevan atau tidak berguna untuk analisis yang

akan dilakukan. Kamus data dalam proses ini menggunakan *library nltk*. Proses ini dapat membantu meningkatkan kualitas data dan meningkatkan efektivitas analisis. Hasil dari proses *filtering* dapat dilihat pada tabel 5.

Tabel 5. Hasil Tahap *Filtering*

Sebelum	Sesudah
saya kira hp saya yang rusak ternyata banyak yg sama keluhannya setelah diperbarui malah setiap kali buka aplikasi tik tok ngelag harus uninstal dan download lagi baru bisa dibuka lagi aplikasinya dan itu mempengaruhi aplikasi lain jadi ngelag semua	hp, rusak, ternyata, keluhannya, diperbarui, kali, buka, aplikasi, tik, tok, ngelag, uninstal, download, dibuka, aplikasinya, mempengaruhi, aplikasi, ngelag

4.2.4. Stemming

Setelah melakukan *filtering* tahap selanjutnya adalah proses *stemming*. *Stemming* sendiri merupakan proses mengubah kata-kata ke bentuk dasarnya. Dalam *stemming*, kata-kata dipotong untuk menghilangkan kata imbuhan pada akhiran atau awalan sehingga hanya menyisakan bentuk dasar. Hasil dari proses *stemming* dapat dilihat pada tabel 6.

Tabel 6. Hasil Tahap *Stemming*

Sebelum	Sesudah
konten video sama diposting dua kali dengan perbedaannya menghilangkan sekian detik bagian depan video tetapi hasilnya satu bisa tayang satu dihapus kok bisa gitu videonya sama lo yg full bisa tayang yg diedit dihilangkan sebagian dihapus aneh banget min yg ngecek video lagi ada masalah apa sih bukti screenshot sudah saya kirim via apps	konten video diposting kali beda hilang sekian detik video hasil tayang hapus gitu video lo full tayang edit hilang hapus aneh banget min ngecek video sih bukti screenshot kirim via apps

4.3. Pelabelan Data

Data yang didapat dari proses *scraping* data melalui *Data Scraper* dan telah melalui proses *preprocessing*, kemudian tahap selanjutnya adalah pelabelan. Dalam melakukan proses pelabelan dilakukan dengan menggunakan metode *lexicon-based*. Metode *lexicon-based* dilakukan dengan cara membuat suatu kamus *lexicon*. Dimana kamus tersebut akan digunakan untuk mengidentifikasi apakah kata tersebut termasuk sentimen positif atau negatif [13]. Proses pembuatan kamus dalam metode *lexicon* dilakukan dengan menggunakan kode program yang memanfaatkan *library nltk* dengan menggunakan jenis *SentimentIntensityAnalyzer*. Dalam kamus *lexicon* setiap kata sudah dilengkapi dengan bobot kata.

Dimana apabila di akumulatiskan kata-kata tersebut dapat digolongkan menjadi kategori sentimen positif, negatif, dan netral. Dinyatakan sebagai sentimen positif apabila skor > 0 atau bernilai positif, apabila dinyatakan sebagai sentimen negatif nilai skor < 0 atau bernilai negatif, dan apabila dinyatakan sebagai sentimen netral nilai skor = 0 atau bernilai netral.

Hasil dari proses pelabelan dengan menggunakan kamus *lexicon* sangat bergantung dengan bobot kata yang telah tersedia di dalam kamus *lexicon*. Hal tersebut mengakibatkan terdapat beberapa kata yang tidak dapat diidentifikasi dengan tepat. Adapun hasil proses pelabelan dengan menggunakan kamus *lexicon* pada Tabel 7 dibawah.

Tabel 7. Hasil Pelabelan

Kelas Sentimen	Kamus Lexicon
Positif	95
Negatif	200
Netral	785
Total	1080

Selanjutnya dilakukan penghilangan sentimen netral pada hasil pelabelan, hal itu dikarenakan pada penelitian ini hanya berfokus pada sentimen negatif dan positif saja sehingga sentimen netral perlu dihilangkan. Pada Tabel 8 merupakan hasil akhir pelabelan kelas sentimen pada ulasan aplikasi *TikTok*.

Tabel 8. Hasil Akhir Pelabelan

Kelas Sentimen	Kamus Lexicon
Positif	95
Negatif	200
Total	295

Maka hasil akhir pelabelan yang didapatkan adalah sebesar 295 dari 1080 ulasan dengan hasil kelas sentimen positif sebesar 95 dan kelas negatif sebesar 200 ulasan.

4.3. Pembobotan

Pada tahap pembobotan kata dalam penelitian ini dilakukan dengan mengimplementasikan algoritma *term frequency-inverse document frequency* (TF-IDF). TF-IDF merupakan suatu metode yang digunakan untuk memberikan pembobotan terhadap keterhubungan suatu kata dengan dokumen. Perhitungan *term frequency* kata yang sering muncul dianggap mewakili sebuah dokumen. Sedangkan proses *term frequency inverse document frequency* semakin sering kata muncul dalam banyak dokumen maka kata tersebut tidak dianggap mewakili sebuah dokumen karena bersifat umum. Proses ini dimulai dengan melakukan perhitungan *term frequency* (TF) kemunculan setiap kata pada setiap kalimat yang ada pada sebuah dokumen dibagi dengan jumlah total kata yang ada pada suatu kalimat. Kemudian menentukan nilai IDF dengan cara melakukan perhitungan nilai log dari hasil *document frequency* (DF) yang telah diproses oleh sistem. Tahap selanjutnya adalah melakukan perhitungan TF-IDF.

4.4. Pembagian Data

Setelah dilakukan proses pelabelan dan pembobotan, maka selanjutnya dilakukan tahap pembagian data, dimana data dibagi menjadi beberapa bagian yaitu data latih (*training*) dan data uji (*testing*). Data latih (*training*) digunakan untuk melatih atau membangun model, sementara data uji (*testing*) digunakan untuk menguji model setelah proses pelatihan selesai. Pembagian data latih (*training*) dan data uji (*testing*) dilakukan untuk mengetahui nilai akurasi yang paling tinggi. Dimana pembagian data yang digunakan dari beberapa penelitian terbagi menjadi beberapa skenario yaitu 90:10, 80:20, 70:30, dan 60:40.

Pada pembagian data latih (*training*) dan data uji (*testing*) diketahui pada percobaan 1 dengan 90% data latih dan 10% data uji. Percobaan 2 dengan 80% data latih dan 20% data uji. Percobaan 3 dengan 70% data latih dan 30% data uji. Percobaan 4 dengan 60% data latih dan 40% data uji.

4.5. Implementasi Pemodelan dengan Naïve Bayes

Setelah dilakukan pembagian data latih dan data uji sesuai dengan skenario yang dilakukan, maka tahap selanjutnya adalah tahap klasifikasi dengan menggunakan metode *naïve bayes*. Proses klasifikasi menggunakan *naïve bayes* memanfaatkan *library sklearn* dengan menggunakan jenis *multinomial naïve bayes*. Kode program *naïve_bayes = MultinomialNB()* berfungsi untuk mendefinisikan variabel *multinomialNB()* dengan nama *naïve_bayes*.

4.6. Analisis dan Hasil

Tahapan hasil uji evaluasi pada penelitian ini menggunakan proses *confusion* matrik yang memanfaatkan *library sklearn* dengan menggunakan *sklearn.metrics* dengan tujuan untuk mengetahui seberapa baik tingkat akurasi pada proses klasifikasi yang telah diproses oleh sistem. Pada penelitian ini, tahapan ini dilakukan sebanyak empat kali percobaan dengan menggunakan data latih (*training*) dan data uji (*testing*) masing-masing sebanyak 90:10, 80:20, 70:30, dan 60:40. Dalam pengimplementasiannya, perhitungan akurasi, presisi, *recall*, dan *f1-score* juga digunakan dalam penelitian ini dengan menggunakan rangkaian kode program. Salah satu hasil skenario dari penerapan dengan model *naïve bayes* dapat dilihat pada Gambar dibawah.

```

Akurasi Model Naive Bayes : 0.7966101694915254
Confusion Matriks:
[[36  3]
 [ 9 11]]
=====
Laporan Klasifikasi :
           precision    recall  f1-score   support

negatif    0.80      0.92      0.86      39
positif    0.79      0.55      0.65      20

accuracy   0.80      0.80      0.80      59
macro avg  0.79      0.74      0.75      59
weighted avg 0.80      0.80      0.79      59
    
```

Gambar 2. Hasil Klasifikasi Model *Naive Bayes*

Hasil data uji evaluasi yang dilakukan dapat dilihat pada Tabel 9 untuk metode *naive bayes* sebagai berikut

Tabel 9. Hasil Uji Evaluasi

Ket.	Percobaan 1	Percobaan 2	Percobaan 3	Percobaan 4
Data Training	90%	80%	70%	60%
Data Testing	10%	20%	30%	40%
True Positive	3	11	14	16
True Negative	20	36	56	74
False Positive	0	3	4	8
False Negative	7	9	15	20
Precision	87%	79%	78%	73%
Recall	65%	74%	71%	67%
F1-Score	66%	75%	73%	69%
Accuracy	77%	80%	79%	76%

Pada tabel diatas, menunjukkan bahwa setelah melakukan sebanyak empat kali percobaan skenario yang memiliki nilai akurasi tertinggi terjadi pada percobaan ke dua dengan perbandingan data *training* 80% dan data *testing* 20% dengan nilai akurasi sebesar 80%.

5. KESIMPULAN DAN SARAN

Data yang telah dikumpulkan dari ulasan penggunaan aplikasi *TikTok* pada *website google play* melalui proses *scraping* kemudian diklasifikasi. Proses klasifikasi diawali dengan *preprocessing*, dilanjutkan dengan pelabelan menggunakan metode *lexicon-based*. Dari proses pelabelan yang dilakukan menghasilkan kelas sentimen positif sebanyak 95 ulasan, kelas sentimen negatif sebanyak 200 dan kelas sentimen netral sebanyak 785 dari 1080 ulasan. Selanjutnya proses mengubah data teks menjadi nilai numerik menggunakan TF-IDF. Proses ini dimulai dengan melakukan perhitungan *term frequency* (TF).

Kemudian menentukan nilai *invers document frequency* (IDF) dengan cara melakukan perhitungan

nilai log dari hasil *document frequency* (DF). Tahap selanjutnya adalah melakukan perhitungan TF-IDF dengan melakukan perkalian nilai *term frequency* (TF) dan nilai *invers document frequency* (IDF). Berdasarkan proses klasifikasi dengan menggunakan algoritma *Naive Bayes* yang dilakukan melalui skenario data latih 80 % dan data uji 20 % menunjukkan nilai akurasi sebesar 80 % dengan *precision* sebesar 79%, *recall* sebesar 74% dan *F1-Score* sebesar 75%.

Saran untuk penelitian selanjutnya diharapkan dapat melakukan pengambilan data dengan rentang yang lebih lama dengan tujuan menunjang keakuratan dari penelitian itu sendiri. Pada proses pengolahan data bagian pelabelan, pada penelitian ini dilakukan dengan menggunakan kamus *lexicon*. Sedangkan pada kamus tersebut tidak semua memiliki kata beserta bobot yang sesuai dengan data yang dimiliki sehingga perlu dilakukan perbaikan dengan mengguna tambahan tools lainnya.

DAFTAR PUSTAKA

- [1] R. Apriani *et al.*, “Analisis Sentimen dengan *Naive Bayes* Terhadap Komentar Aplikasi Tokopedia,” *J. Rekayasa Teknol. Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, [Online]. Available: <https://rekayasa.nusaputra.ac.id/article/view/86>
- [2] Wikipedia, “TikTok. (2023, September 29). Di Wikipedia, Ensiklopedia Bebas. Diakses pada 10:19, September 29, 2023, dari,” <https://id.wikipedia.org/w/index.php?title=TikTok&oldid=24361164>, 2023.
- [3] Wikipedia, “Google Play. (2023, November 10). Di Wikipedia, Ensiklopedia Bebas. Diakses pada 18:00, November 10, 2023, dari,” https://id.wikipedia.org/w/index.php?title=Google_Play&oldid=24762382, 2023.
- [4] V. A. Flores, L. Jasa, and L. Linawati, “Analisis Sentimen untuk Mengetahui Kelemahan dan Kelebihan Pesaing Bisnis Rumah Makan Berdasarkan Komentar Positif dan Negatif di Instagram,” *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 1, p. 49, 2020, doi: 10.24843/mite.2020.v19i01.p07.
- [5] W. Parasati, F. Abdurrachman Bachtiar, and N. Y. Setiawan, “Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode *Naive Bayes Classifier*,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4, pp. 1090–1099, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] A. N. Ardianti and M. A. Widiartanto, “Pengaruh Online Customer Review dan Online Customer Rating terhadap Keputusan Pembelian melalui Marketplace Shopee .,” *J. Ilmu Adm. Bisnis*, pp. 1–11, 2019.
- [7] A. T. J. Harjanta, “Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining,” vol. 1, pp. 1–9, 2015.

- [8] T. Imandasari, E. Irawan, A. P. Windarto, and A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 750, 2019, doi: 10.30645/senaris.v1i0.81.
- [9] Y. A. Wijaya, N. Suarna, Iin, R. Hamonangan, and R. Nining, "Comparison of machine learning algorithm for Santander dataset," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012032, 2021, doi: 10.1088/1757-899x/1088/1/012032.
- [10] S. Rahayu, J. J. Jaya, H. M. Mahmud, and F. S. Nugraha, "Algoritma Naive Bayes Classifier Untuk Memprediksi Gejala Autism Spectrum Disorders Pada Anak-Anak," *Semin. Nas. Rekayasa dan Teknol.*, vol. 27, no. November, pp. 62–74, 2019.
- [11] M. Arvyantomo and N. Ratama, "Analisis Sentimen Masyarakat Indonesia terhadap Invasi Russia di Ukraina menggunakan Metode Naive Bayes pada Media Sosial Facebook," *Log. J. Ilmu Komput. ...*, vol. 1, no. 4, pp. 705–717, 2023, [Online]. Available: <https://www.journal.mediapublikasi.id/index.php/logic/article/view/2142>
- [12] R. O. Felani, "Analisis Prilaku Pengguna e-learning menggunakan Algoritma K-Means Clustering," *Jusikom J. Sist. Komput. Musirawas*, vol. 7, no. 1, pp. 61–73, 2022, doi: 10.32767/jusikom.v7i1.1538.
- [13] A. Kurniawan, Indriati, and S. Adinugroho, "Analisis Sentimen Opini Film Menggunakan Metode Naive Bayes dan Lexicon Based Features," *J. Pengemb. Teknol. Inf. dan Komput.*, vol. 3, no. 9, pp. 8335–8342, 2019, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6153>