

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MEDIA SOSIAL INSTAGRAM PADA SITUS GOOGLE PLAY STORE MENGGUNAKAN *NAÏVE BAYES CLASSIFIER*

Reza Zulfiqri¹, Betha Nurina Sari², Tesa Nur Padilah³

^{1,2,3}Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361; Telp. (0267) 641177

Received: 14 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Scraping;

Instagram;

Analisis Sentimen;

Google Play Store;

Correspondent Email:

rezazulfiqri07@gmail.com

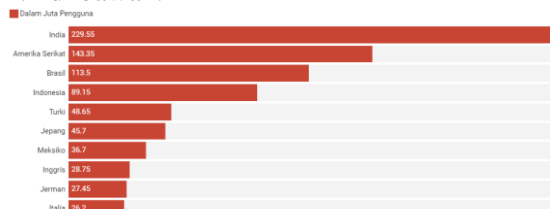
Abstrak. Instagram merupakan salah satu media sosial paling populer di seluruh dunia, dengan lebih dari satu miliar pengguna aktif setiap bulannya. Semakin berkembangnya teknologi informasi pada era digital seperti sekarang membuat opini atau sentimen masyarakat banyak dibagikan dalam kolom komentar atau ulasan. Oleh karena itu perlu dilakukan analisis terhadap ulasan aplikasi media sosial instagram untuk melihat opini masyarakat mengenai aplikasi tersebut. Data ulasan diperoleh dari situs web Google Play yang diambil menggunakan metode *scraping*. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi media sosial Instagram yang terdapat pada situs Google Play Store untuk mengetahui dan mengklasifikasikan komentar baik dan buruk. Sentimen pengguna dapat memberikan wawasan yang berharga bagi pengembang aplikasi dalam memahami kepuasan pengguna, mengevaluasi fitur yang ada, serta meningkatkan kualitas aplikasi. Pada pengujian ini menggunakan 1000 data terdiri dari 80% data *training* dan 20% data *testing*. Hasil pengujian akurasi, penelitian ini menggunakan *confusion matrix*, dalam penelitian ini algoritma *Naïve Bayes classifier* dengan akurasi yang didapatkan sebesar 89%.

Abstract. Instagram is one of the most popular social media platforms worldwide, with over one billion active users each month. The advancement of information technology in today's digital era has led to a surge in public opinions or sentiments shared in comment sections or reviews. Therefore, it is necessary to analyze reviews of the Instagram social media application to understand public opinions about the app. The review data was obtained from the Google Play website using the scraping method. This research aims to perform sentiment analysis on user reviews of the Instagram social media application found on the Google Play Store to identify and classify positive and negative comments. User sentiments can provide valuable insights for application developers to understand user satisfaction, evaluate existing features, and improve the application's quality. In this testing, 1000 data samples were used, consisting of 80% training data and 20% testing data. The accuracy of the test results was measured using a confusion matrix. In this study, the Naïve Bayes classifier algorithm achieved an accuracy of 89%.

1. PENDAHULUAN

Kebutuhan akan berbagai momen, menjalin interaksi dan mendapatkan inspirasi dengan mudah membuat tim pengembang atau perusahaan teknologi menciptakan inovasi untuk membangun berbagai platform dan aplikasi media sosial. Perkembangan teknologi informasi yang sangat pesat mendorong inovasi dalam menciptakan platform-platform dan media sosial yang sudah dibangun pada saat ini. Dengan terciptanya berbagai platform dan media sosial, pengguna dapat berbagi, berinteraksi, dan mendapatkan informasi tentang banyak hal dengan mudah.[1]

Instagram merupakan salah satu inovasi yang diciptakan oleh perusahaan teknologi Facebook. Nama instagram kependekan dari kata “instan-telegram” yang merupakan layanan jejaring sosial berbasis fotografi. Jejaring sosial ini diresmikan pada tanggal 6 Oktober 2010 oleh Kevin Systrom dan Mike Krieger yang mampu menjaring 25 ribu pengguna di hari pertama. Data berdasarkan CNBC Indonesia tahun 2023 tercatat bahwa pada awal tahun 2023 Indonesia menjadi negara dengan jumlah pengguna instagram terbanyak ke-4 di dunia, yaitu sebanyak 89,15 juta pengguna, seperti yang terdapat pada gambar 1.1 di bawah.



Gambar 1.1 Negara dengan jumlah pengguna instagram terbanyak di dunia (sumber : *we are social report*)

Seiring dengan popularitasnya, banyak pengguna yang memberikan ulasan atau *review* tentang pengalaman mereka menggunakan aplikasi Instagram di situs Google Play Store. Ulasan dari setiap pengguna dapat berpengaruh terhadap pertimbangan dan keberlangsungan penggunaan suatu aplikasi, termasuk pada aplikasi media sosial Instagram. Pengguna dapat mengutarakan pandangan terhadap layanan aplikasi Instagram baik itu berupa pujian atau masukan. Banyaknya pengguna Instagram saat ini mengakibatkan data ulasan semakin banyak. Dengan banyaknya data

ulasan, kita tidak mungkin untuk mengolahnya secara manual. Maka dari itu diperlukan metode untuk mengklasifikasikan ulasan pengguna terhadap aplikasi tersebut secara otomatis, apakah ulasan positif lebih banyak dibanding ulasan *negative* atau sebaliknya [2]

Berdasarkan beberapa penelitian di atas, belum ada penelitian yang dilakukan terhadap sentimen user Instagram, maka dari itu peneliti ingin melakukan penelitian terhadap sentimen ulasan user Instagram dengan tujuan untuk mengetahui Bagaimana cara mengklasifikasikan ulasan pengguna aplikasi media sosial Instagram ke dalam kelas sentimen positif dan negatif menggunakan *algoritme naïve bayes classifier* dan Bagaimana gambaran atau informasi umum mengenai data ulasan pengguna aplikasi media sosial Instagram pada situs Google Play Store.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

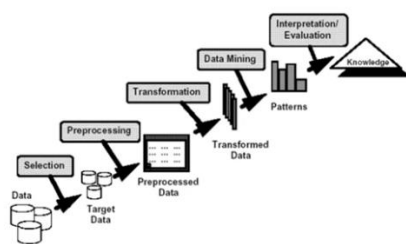
Studi komputasi mengenai perasaan, evaluasi, pandangan, suasana hati, serta emosi individu sering disebut sebagai analisis sentimen atau ekstraksi opini. Analisis sentimen adalah salah satu bidang penelitian yang sangat aktif dalam domain pemrosesan bahasa alami, penambahan data, pengambilan informasi, dan penambahan web [3]

Analisis sentimen, yang juga dikenal sebagai ekstraksi opini, terutama memanfaatkan pemrosesan bahasa alami dan teknik ekstraksi informasi untuk mengidentifikasi dan menganalisis teks [4]. Penilaian terhadap orientasi tertentu dalam teks sangat tergantung pada konteks dan polaritas yang termanifestasi, dan ini bisa mencakup argumen, opini, atau ekspresi sentimen dalam teks.

Menurut analisis sentimen merupakan kumpulan pandangan seseorang tentang suatu peristiwa apapun yang terjadi dalam kehidupan nyata. Sedangkan analisis sentimen media sosial sendiri yaitu teknik atau metode yang berkembang untuk memahami pendapat suatu individu melalui jejaring sosial [5]

2.2 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Databases (KDD) atau Penemuan Pengetahuan dalam Basis Data adalah proses mendapatkan informasi yang berguna, yang sebelumnya tidak diketahui, dari data yang ada. KDD melibatkan serangkaian langkah sistematis yang dilakukan untuk mengungkap pola, hubungan, dan pengetahuan baru dari data yang besar, kompleks, dan seringkali tidak terstruktur [6]. Seperti struktur yang disajikan oleh gambar 2.1.



Gambar 2.1 Tahapan *Knowledge Discovery in Database*
(Sumber: Jiawei Han, 2023)

2.3 Algoritme Naïve Bayes

Naïve Bayes Classifier merupakan suatu metode probabilistik dengan independensi yang kuat diantara fitur-fiturnya dan mengasumsikan bahwa nilai fitur tertentu menentukan sendiri nilai fitur lainnya dengan cara mempertimbangkan variabel kelas [7]. Dalam pendekatan supervised learning, *Naïve Bayes Classifier* dapat melatih data dengan efisien dan merupakan model probabilitas bersyarat.

Naïve Bayes Classifier, mengambil dasar dari *Teorema Bayes* yang memberikan probabilitas kondisional terhadap setiap kelas berdasarkan data yang ada. Akan tetapi, hipotesis yang mendasari NBC adalah asumsi ketidak adaan ketergantungan antar variabel dalam suatu sistem. Walaupun asumsi ini dianggap tidak realistis, metode klasifikasi berdasarkan *Naïve Bayes Classifier* (NBC) mampu memberikan hasil yang memuaskan dalam konteks masalah klasifikasi sebenarnya [8].

2.4 Term Frequency – Inverse Document Frecuency (TF – IDF)

TF – IDF adalah sebuah algoritme transformasi data teks menjadi numerik. TF – IDF merupakan perpaduan perhitungan

antara TF dan IDF. TF (*Term Frequency*) adalah rasio dari jumlah suatu kata pada kalimat yang dibandingkan dengan panjang dari kalimat tersebut. Sedangkan IDF (*Inverse Document Frequency*) dari setiap kata adalah log atau ratio berdasarkan total dokumen dengan jumlah dari dokumen tertentu yang dimana kata tersebut ada. Jadi TF – IDF adalah gabungan dari TF dan IDF yang dimana keduanya dapat menutupi kekurangan untuk membuat prediksi menjadi relevan [7]. Persamaan untuk perhitungan TF – IDF menurut [9] dapat dilihat di bawah ini:

$$1. \text{ TF} \quad TF_{t,d} = F(t,d) \quad (2.2)$$

$$2. \text{ IDF} \quad IDF_t = \log \frac{N}{D_{ft}} \quad (2.3)$$

$$3. \text{ TF-IDF} \quad wt = TF_{t,d} \times IDF_t^2 \quad (2.4)$$

Gambar 2.2 Rumus TF-IDF

Keterangan

t : Fitur

d : Dokumen

D_{ft} : Dokumen yang mengandung fitur

N : Banyaknya Dokumen

2.5 Confusion Matrix

Confusin matrix dilakukan untuk melihat kinerja dari sebuah metode *Classifier* atau *Supervised Learning* dimana setiap kolom yang terdapat pada matriks mewakili sebuah kelas yang akan dilakukan sebuah proses perdisksi [10]. Klasifikasi memerlukan sebuah perhitungan dari beberapa data yang terdapat pada tabel matriks dengan menggunakan 4 (empat) istilah yaitu *True Positive*, *False Positive*, *True Negative*, dan *False Negative* untuk dijadikan sebagai representasi dari hasil klasifikasi [11]. Dalam pengujian evaluasi ini, menyatakan bahwa hasil yang diperoleh akan mencakup nilai *accuracy*, *precision*, dan *recall*.

2.6 Python

Python adalah salah satu bahasa pemrograman yang memiliki tujuan umum terbaik yang bisa digunakan diberbagai sistem operasi saat ini. *Python* juga merupakan bahasa pemrograman dengan tipe *interpreted language* yang dimana code tidak akan diubah menjadi sebuah code yang dapat dipahami oleh komputer sebelum dijalankan, melainkan perubahan

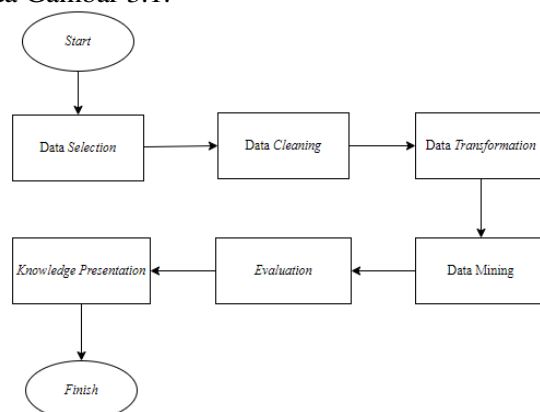
tersebut terjadi saat *runtime* [12]. *Python* sendiri diciptakan oleh Guido Van Rossum. Aplikasi yang bisa dikembangkan dari bahasa pemrograman ini sangat beragam salah satunya dibidang *Scientific* dan *numeric* yang memiliki *library* cukup banyak seperti *Spicy*, *Pandas*, *Numpy* (Python Software Foundation, 2021).

2.7 Instagram

Instagram adalah platform media sosial yang memungkinkan pengguna untuk berbagi foto, video, dan cerita dalam format visual. Diluncurkan pada tahun 2010, Instagram awalnya fokus pada berbagi foto dengan filter artistik untuk meningkatkan penampilan gambar. Namun, seiring berjalannya waktu, platform ini berkembang menjadi salah satu media sosial terbesar di dunia dengan lebih dari satu miliar pengguna aktif bulanan pada tahun 2023[13]

3. METODE PENELITIAN

Tahapan-tahapan yang diimplementasikan pada penelitian ini mengacu pada metodologi KDD yang terdiri dari 5 tahap yaitu *Data Selection*, *Data PreProcessing*, *Data Transformation*, *Data Mining* dan *Evaluation*. Berikut adalah rancangan penelitian analisis sentimen ulasan pengguna aplikasi media sosial instagram pada situs *google play store* menggunakan *naïve bayes classifier* terdapat pada Gambar 3.1.



Gambar 3.1 Diagram Alur Metode Penelitian

Dari alur metode penelitian di atas, berikut penjelasan dari tiap tahapan tersebut.

3.1 Data Selection

Tahap ini merupakan tahap pertama dalam penelitian ini. Teknik pengumpulan data yang peneliti gunakan adalah dengan metode *scraping* dengan menggunakan API Instagram. *Data scraping* adalah proses mengambil data dari sumber lain secara otomatis menggunakan program atau *algoritme* khusus. Tujuan dari proses ini adalah untuk mengambil data yang terstruktur dan dapat diakses melalui internet atau sumber elektronik lainnya dan mengubahnya menjadi format yang dapat digunakan untuk keperluan analisis lainnya.

3.2 Data Preprocessing/Cleaning

Pada tahap ini dilakukan untuk menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (*discovery*) akan dilakukan. Tahapan ini berkaitan dengan pemilihan data yang akan digunakan dalam proses KDD [14]. Selection atau pemilihan data harus didasarkan pada tujuan analisis dan jenis data yang diperlukan.

3.2.1 Cleaning

Cleaning adalah proses persiapan data yang melibatkan serangkaian tahapan untuk membersihkan, mengubah, dan menyesuaikan data teks mentah menjadi format yang lebih siap digunakan untuk analisis dan pemrosesan selanjutnya.

3.2.2 Case Folding

Proses ini adalah proses mengubah seluruh huruf menjadi huruf kecil. Pada proses ini karakter-karakter ‘A’-‘Z’ yang terdapat pada data diubah ke dalam karakter ‘a’-‘z’.

3.2.3 Stopword Removal

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopword* dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan *stopword* yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya.

3.2.4 Tokenizing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian dianalisis. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Didalam

NLP, token diartikan sebagai “kata” meskipun *tokenize* juga dapat dilakukan pada paragraf maupun kalimat.

3.2.5 Stemming

Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya.

3.3 Data Transformation

Setelah tahap *pre-processing* selesai, dilanjutkan dengan tahap transformasi atau merubah teks menjadi suatu data yang memiliki tipe numerik sebagai representasi dari suatu dokumen. Penerapan pembobotan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) dilakukan dengan menghitung bobot untuk setiap kata. Metode ini dapat menghasilkan skor frekuensi kata yang penting atau kata-kata dengan tingkat kemunculan yang tinggi pada dokumen yang sedang dianalisis.

3.4 Data Mining

Tahap berikutnya adalah tahap data *mining* di mana data mentah diubah menjadi informasi yang bermanfaat. Proses penambangan data melibatkan eksplorasi dan analisis untuk mengidentifikasi pola yang berguna dan informatif. Dalam konteks ini, metode klasifikasi digunakan untuk mengelompokkan kalimat sentimen ke dalam kategori positif atau negatif. *Algoritme* yang digunakan adalah *Naïve Bayes Classifier*, yang merujuk pada Teorema *Bayes* dan memiliki acuan untuk melakukan klasifikasi.

3.5 Evaluation

Di tahap ini, pengujian dilakukan menggunakan *Confusion Matrix* untuk mengevaluasi performa klasifikasi. Hasilnya akan berupa nilai akurasi (*accuracy*), presisi (*precision*), dan *recall*. Dengan menggunakan nilai-nilai ini, dapat dievaluasi seberapa baik dan akurat proses klasifikasi tersebut. Selain itu, hasil klasifikasi ulasan sentimen positif dan negatif juga akan ditampilkan.

4. HASIL DAN PEMBAHASAN

Berikut merupakan hasil dari penelitian ulasan aplikasi Instagram di Playstore menggunakan algoritma *Naïve Bayes Classifier*. Penelitian ini melalui tahapan-tahapan yang meliputi data selection,

preprocessing, transformation, data mining, dan evaluation.

4.1 Data Selection

Tahap ini merupakan tahap pertama dalam penelitian ini. Teknik pengumpulan data yang peneliti gunakan adalah dengan metode *scraping* dengan menggunakan API Instagram. Data *scraping* adalah proses mengambil data dari sumber lain secara otomatis menggunakan program atau *algoritme* khusus. Tujuan dari proses ini adalah untuk mengambil data yang terstruktur dan dapat diakses melalui internet atau sumber elektronik lainnya dan mengubahnya menjadi format yang dapat digunakan untuk keperluan analisis lainnya.

4.2 Preprocessing

Pada tahap ini dilakukan untuk menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (*discovery*) akan dilakukan. Tahapan ini berkaitan dengan pemilihan data yang akan digunakan dalam proses KDD. Selection atau pemilihan data harus didasarkan pada tujuan analisis dan jenis data yang diperlukan.

4.2.1 Cleaning

Cleaning adalah proses persiapan data yang melibatkan serangkaian tahapan untuk membersihkan, mengubah, dan menyesuaikan data teks mentah menjadi format yang lebih siap digunakan untuk analisis dan pemrosesan selanjutnya.

| index | username | score | at | content | label | text_clean |
|-------|----------------------|-------|---------------------|---|---------|---|
| 0 | indrii_sori | 5 | 2023-06-23 05:40:02 | apik ini bagus banget! beda sama yg updatein itu | Positif | apik ini bagus banget! beda sama yg updatein itu |
| 1 | Riko Irawan | 5 | 2023-06-23 05:43:40 | Aplikasi ngi menarik | Positif | aplikasi ngi menarik |
| 2 | guyon team | 5 | 2023-06-23 05:43:27 | EXCITED AND NICE EXPERIENCE | Positif | excited and nice experience |
| 3 | Almasdi Bida | 1 | 2023-06-23 05:43:23 | hi apik bagus lagi dan banyak nya apik yg yang paling sering ku pakai | Negatif | hi apik bagus lagi dan banyak nya apik yg yang paling sering ku pakai |
| 4 | lrian Ratnaga | 5 | 2023-06-23 05:41:02 | udah saya perbaiki kok ehnya ngk keluar, ehk orange itu bu? tolong diperbaiki lagi | Positif | udah saya perbaiki kok ehnya ngk keluar, ehk orange itu bu? tolong diperbaiki lagi |
| 5 | Rendi Hamdani | 1 | 2023-06-23 05:40:08 | sebelum di update versi terbaru? bukannya makin baik, malah semakin bu? tolong diperbaiki ke versi semula lagi!! parah Instagram | Negatif | sebelum di update versi terbaru? bukannya makin baik, malah semakin bu? tolong diperbaiki ke versi semula lagi!! parah Instagram |
| 6 | Hasbi Putra | 5 | 2023-06-23 05:39:36 | soal nya bagus | Positif | soal nya bagus |
| 7 | Mulana Akmalul Iqbal | 2 | 2023-06-23 05:39:03 | Setelah di update, posting yg muncul malah posting yg sudah dihapus, yg baru malah ketembur, & kadang ga ada tanda ceklis lagi ya | Negatif | Setelah di update, posting yg muncul malah posting yg sudah dihapus, yg baru malah ketembur, & kadang ga ada tanda ceklis lagi ya |
| 8 | adri andia | 5 | 2023-06-23 05:39:37 | keren | Positif | keren |
| 9 | Nasib Ananda | 5 | 2023-06-23 05:39:14 | Good job | Positif | good job |

Gambar 3.2 Proses *Cleaning*

4.2.2 Case Folding

Proses ini adalah proses mengubah seluruh huruf menjadi huruf kecil. Pada proses ini karakter-karakter ‘A’-‘Z’ yang terdapat pada data diubah ke dalam karakter ‘a’-‘z’ seperti yang terlihat pada gambar 3.3.

| index | username | score | at | content | Label | text_clean | text_StopWord |
|-------|---------------------|-------|---------------------|---|---------|--|---|
| 0 | akrida wati | 5 | 2023-06-05 05:40:02 | apk ini bagus banget👍 beda sama yg updetan lu👍👍👍 | Positif | apk ini bagus banget👍 beda sama yg updetan lu👍👍👍 | apk bagus banget👍 beda yg updetan lu👍👍👍 |
| 1 | Rika Irawan Ginting | 5 | 2023-06-05 05:43:48 | Aplikasi sgt menarik | Positif | aplikasi sgt menarik | aplikasi sgt menarik |
| 2 | gibler team | 5 | 2023-06-05 05:43:27 | EXCITED AND NICE EXPERIENCE | Positif | excited and nice experience | excited and nice experience |
| 3 | Ahmad Bilqis | 1 | 2023-06-05 05:43:23 | Ni apk bagus bgt dari banyak nya apk yg paling sering ku pakai | Negatif | ni apk bagus bgt dari banyak nya apk yg paling sering ku pakai | ni apk bagus bgt nya apk yg ku pakai |
| 4 | Intan Rahayu | 5 | 2023-06-05 05:41:02 | Udah saya perbarui kok efeknya ngk keluar2, ehk orange lu ku773 yang diperbaiki lagi | Positif | udah saya perbarui kok efeknya ngk keluar2, ehk orange lu ku773 yang diperbaiki lagi | udah perbarui efeknya ngk keluar2, ehk orange lu773 yang diperbaiki lagi |
| 5 | Randi Nuzumandi | 1 | 2023-06-05 05:40:08 | Sekali di update Versi Terbaru Bukannya Baik, Malah semakin Buruk. Tolong diperbaiki Ku Versi Semula lagi!! Parah Instagram 🤬🤬🤬 | Negatif | sekal di update versi terbaru bukannya malah baik, malah semakin buruk... tolong diperbaiki ku versi semula lagi!! Parah Instagram 🤬🤬🤬 | update versi terbaru baik, buruk... tolong diperbaiki versi!! Parah Instagram 🤬🤬🤬 |
| 6 | Hadi Putra | 5 | 2023-06-05 05:38:06 | sual nya bagus | Positif | sual nya bagus | nya bagus |

Gambar 3.3 Proses Case Folding

4.2.3 Stopword Removal

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopword* dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan *stopword* yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya.

| index | username | score | at | content | Label | text_clean | text_StopWord |
|-------|---------------------|-------|---------------------|---|---------|--|---|
| 0 | akrida wati | 5 | 2023-06-05 05:40:02 | apk ini bagus banget👍 beda sama yg updetan lu👍👍👍 | Positif | apk ini bagus banget👍 beda sama yg updetan lu👍👍👍 | apk bagus banget👍 beda yg updetan lu👍👍👍 |
| 1 | Rika Irawan Ginting | 5 | 2023-06-05 05:43:48 | Aplikasi sgt menarik | Positif | aplikasi sgt menarik | aplikasi sgt menarik |
| 2 | gibler team | 5 | 2023-06-05 05:43:27 | EXCITED AND NICE EXPERIENCE | Positif | excited and nice experience | excited and nice experience |
| 3 | Ahmad Bilqis | 1 | 2023-06-05 05:43:23 | Ni apk bagus bgt dari banyak nya apk yg paling sering ku pakai | Negatif | ni apk bagus bgt dari banyak nya apk yg paling sering ku pakai | ni apk bagus bgt nya apk yg ku pakai |
| 4 | Intan Rahayu | 5 | 2023-06-05 05:41:02 | Udah saya perbarui kok efeknya ngk keluar2, ehk orange lu ku773 yang diperbaiki lagi | Positif | udah saya perbarui kok efeknya ngk keluar2, ehk orange lu ku773 yang diperbaiki lagi | udah perbarui efeknya ngk keluar2, ehk orange lu773 yang diperbaiki lagi |
| 5 | Randi Nuzumandi | 1 | 2023-06-05 05:40:08 | Sekali di update Versi Terbaru Bukannya Baik, Malah semakin Buruk. Tolong diperbaiki Ku Versi Semula lagi!! Parah Instagram 🤬🤬🤬 | Negatif | sekal di update versi terbaru bukannya malah baik, malah semakin buruk... tolong diperbaiki ku versi semula lagi!! Parah Instagram 🤬🤬🤬 | update versi terbaru baik, buruk... tolong diperbaiki versi!! Parah Instagram 🤬🤬🤬 |
| 6 | Hadi Putra | 5 | 2023-06-05 05:38:06 | sual nya bagus | Positif | sual nya bagus | nya bagus |

Gambar 3.4 Proses Stopword Removal

4.2.4 Tokenizing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian dianalisis. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Didalam NLP, token diartikan sebagai “kata” meskipun *tokenize* juga dapat dilakukan pada paragraf maupun kalimat. Berikut hasil *tokenizing* yang telah dibuat dapat dilihat pada gambar 3.5.

| index | username | score | at | content | Label | text_clean | text_tokens |
|-------|---------------------|-------|---------------------|--|---------|--|--|
| 0 | akrida wati | 5 | 2023-06-05 05:40:02 | apk ini bagus banget👍 beda sama yg updetan lu👍👍👍 | Positif | apk ini bagus banget👍 beda sama yg updetan lu👍👍👍 | apk bagus banget👍 beda yg updetan lu👍👍👍 |
| 1 | Rika Irawan Ginting | 5 | 2023-06-05 05:43:48 | Aplikasi sgt menarik | Positif | aplikasi sgt menarik | aplikasi sgt menarik |
| 2 | gibler team | 5 | 2023-06-05 05:43:27 | EXCITED AND NICE EXPERIENCE | Positif | excited and nice experience | excited and nice experience |
| 3 | Ahmad Bilqis | 1 | 2023-06-05 05:43:23 | Ni apk bagus bgt dari banyak nya apk yg paling sering ku pakai | Negatif | ni apk bagus bgt dari banyak nya apk yg paling sering ku pakai | ni apk bagus bgt nya apk yg ku pakai |
| 4 | Intan Rahayu | 5 | 2023-06-05 05:41:02 | Udah saya perbarui kok efeknya ngk keluar2, ehk orange lu ku773 yang diperbaiki lagi | Positif | udah saya perbarui kok efeknya ngk keluar2, ehk orange lu ku773 yang diperbaiki lagi | udah perbarui efeknya ngk keluar2, ehk orange lu773 yang diperbaiki lagi |

Gambar 3.5 Proses Tokenizing

4.2.5 Stemming

Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya. Berikut hasil *stemming* yang telah dilakukan seperti terlihat pada gambar 3.6.

1 : apk : apk
 2 : bagus : bagus
 3 : banget👍 : banget
 4 : beda : beda
 5 : yg : yg
 6 : updetan : updetan
 7 : itu🤬🤬🤬🤬 : itu
 8 : aplikasi : aplikasi
 9 : sgt : sgt
 10 : menarik : tarik
 11 : excited : excited
 12 : and : and
 13 : nice : nice
 14 : experience : experience
 15 : ni : ni
 16 : bgt : bgt
 17 : nya : nya
 18 : ig : ig
 19 : ku : ku
 20 : pakai : pakai
 21 : udah : udah
 22 : perbarui : baru
 23 : efeknya : efek
 24 : nggk : nggk
 25 : keluar : keluar

Gambar 3.6 Proses Stemming

4.3 Transformation.

Setelah penyelesaian tahap pre-processing, langkah selanjutnya adalah tahap transformasi, di mana teks diubah menjadi data yang memiliki representasi numerik. Pembobotan diterapkan dengan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency), dimana bobot untuk setiap kata dihitung. Pendekatan ini menghasilkan skor yang mencerminkan frekuensi kata yang signifikan atau kata-kata yang sering muncul dalam dokumen yang sedang dianalisis.

```
[31] from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer()
tfidf_train = tfidf_vectorizer.fit_transform(X_train)
tfidf_test = tfidf_vectorizer.transform(X_test)

[32] print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)

(736,)
(736,)
(184,)
(184,)

[34] X_train = vectorizer.transform(X_train)
X_test = vectorizer.transform(X_test)

[35] from sklearn.naive_bayes import MultinomialNB
nb = MultinomialNB()
nb.fit(tfidf_train, y_train)

MultinomialNB()

[36] X_train.toarray()

array([[0, 0, ..., 0, 0, 0],
       [0, 0, ..., 0, 0, 0],
       [0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, ..., 0, 0, 0],
       [0, 0, ..., 0, 0, 0],
       [0, 0, ..., 0, 0, 0]])
```


dilakukan melalui beberapa tahapan penting, yaitu *data selection*, *data preprocessing*, *data transformation*, *data mining*, dan *evaluation*. Tahap pertama, *data selection*, melibatkan pengumpulan data melalui metode scraping menggunakan API Instagram, yang bertujuan untuk mengumpulkan data terstruktur dari internet. Setelah data terkumpul, tahap *data preprocessing* dilakukan untuk mempersiapkan data tersebut agar siap dianalisis. Langkah-langkah dalam *preprocessing* meliputi cleaning (membersihkan dan mengubah data teks mentah menjadi format yang lebih siap), case folding (mengubah semua huruf menjadi huruf kecil), *stopword removal* (menghapus kata-kata umum yang tidak bermakna signifikan), tokenizing (memisahkan teks menjadi token), dan stemming (mengubah kata menjadi bentuk dasar).

Setelah *preprocessing*, tahap *data transformation* dilakukan dengan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengubah teks menjadi representasi numerik, yang memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya. Langkah selanjutnya adalah *data mining*, di mana data yang telah diproses digunakan untuk mengidentifikasi pola atau informasi berguna. Dalam penelitian ini, metode klasifikasi digunakan untuk mengelompokkan sentimen pengguna ke dalam kategori positif atau negatif dengan menggunakan algoritma *Naïve Bayes Classifier*. Data dibagi menjadi 80% *data training* dan 20% *data testing* untuk keperluan pengujian.

Evaluasi kinerja klasifikasi dilakukan menggunakan Confusion Matrix, yang menghasilkan nilai akurasi, presisi, dan recall. Hasil evaluasi menunjukkan bahwa algoritma *Naïve Bayes Classifier* mencapai akurasi sebesar 89%. Presisi untuk kelas negatif adalah 78% dan untuk kelas positif adalah 92%, sedangkan recall untuk kelas negatif adalah 74% dan untuk kelas positif adalah 94%. Hasil ini menunjukkan bahwa algoritma *Naïve Bayes Classifier* mampu mengklasifikasikan sentimen pengguna aplikasi Instagram dengan tingkat akurasi yang tinggi. Selain itu, visualisasi hasil analisis sentimen dilakukan dengan menggunakan *WordCloud*, yang menampilkan

kata-kata yang paling sering muncul dalam ulasan pengguna. Kata-kata yang lebih sering muncul ditampilkan dengan ukuran yang lebih besar dalam *WordCloud*, memberikan gambaran visual tentang kata-kata yang paling relevan dalam ulasan tersebut.

5. KESIMPULAN

Berdasarkan hasil penelitian yang dicoba pada pengklasifikasian analisis sentimen terhadap pengguna aplikasi Instagram menggunakan algoritma *Naïve Bayes Classifier* terdapat beberapa tahapan yang harus dilakukan seperti *data selection*, *data preprocessing*, *data transformasi*, *data mining* dan *evaluation*. Pada pengujian dibagi menjadi 2 yaitu 80% *data training* dan 20% *data testing*. Kemudian pada pengujian akurasi, penelitian ini menggunakan *confusion matrix*, dalam penelitian ini akurasi yang didapatkan untuk metode *Naïve Bayes classifier* menghasilkan nilai *accuracy* sebesar 89%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ibu Betha Nurina Sari, M.Kom., dan Ibu Tesa Nur Padilah, S.Si., M.Sc., selaku Dosen Prodi Informatika Universitas Singaperbangsa Karawang yang telah membimbing selama penyusunan skripsi yang kemudian penulis jadikan jurnal. Kemudian, penulis juga ingin berterima kasih kepada pihak terlibat yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes (Yuma Akbar 1*, Tri Sugiharto 2) Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 115–122, 2023, doi: 10.55338/saintek.v4i3.1368.
- [2] D. Surya Sayogo, B. Irawan, and A. Bahtiar, "Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naïve Bayes," 2023.
- [3] R. Fajar, S. Program, P. Rekayasa, N. Lunak, and R. Bengkalis, "Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," vol. 3, no. 1.

- [4] A. Wildan Attabi', L. Muflikhah, and M. A. Fauzi, "Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada Twitter Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [5] A. A. Fani, P. Sudarmaningtyas, and V. Nurcahyawati, "Analisis Sentimen Review Pelanggan Pada Layanan Indihome Menggunakan Algoritma Naïve Bayes Classifier," 2019.
- [6] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *Jurnal CoreIT*, vol. 7, no. 1, 2021.
- [7] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *JURNAL TEKNIK INFORMATIKA*, vol. 11, no. 2, pp. 149–164, Nov. 2018, doi: 10.15408/jti.v11i2.8623.
- [8] R. Apriani *et al.*, "Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia," 2019.
- [9] A. Bijaksana, P. Negara, H. Muhandi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes Dan Seleksi Fitur Information Gain Sentiment Analysis On Airlines Using Naïve Bayes Method And Feature Selection Information Gain," vol. 7, no. 3, pp. 599–606, 2020, doi: 10.25126/jtiik.202071947.
- [10] F. Satria, W. Hayuhardhika, N. Putra, and B. T. Hanggara, "Analisis Sentimen Mengenai Kepuasan Pengguna terhadap Pelayanan Transportasi Minibus PT Mahkota Tria Wisata dengan menggunakan Metode *Support Vector Machine*," 2023. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] F. Solihin, S. Awaliyah, A. Muid, and A. Shofa, "Pemanfaatan Twitter Sebagai Media Penyebaran Informasi Oleh Dinas Komunikasi dan Informatika," *Jurnal Pendidikan Ilmu Pengetahuan Sosial (JPIPS)*, vol. 1, no. 13, pp. 52–58, 2021, [Online]. Available: <http://e-journal.upr.ac.id/index.php/JP-IPS>
- [12] A. V. Sudiantoro and E. Zuliarso, *analisis sentimen twitter menggunakan text mining dengan algoritma naïve bayes classifier*. 2018.
- [13] E. Putri Nirwandani and R. Cahya Wihandika, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode *Modified Term Frequency Scheme* Dan *Naïve Bayes*," 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [14] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis Sentimen Ulasan Pengguna Aplikasi Samsat Digiital Nasional (*Signal*) Dengan Menggunakan Metode *Naïve Bayes Classifier*," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4073.