

Analisis Sentimen Ulasan *User* Terhadap Aplikasi Tix-Id Menggunakan Metode *Naïve Bayes* Pada *Google Play Store*

Java Khitamasari^{1*}, Isa Faqihuddin Hanif²

¹Program Studi Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Indonesia

Email: ¹javakhitamasari5@gmail.com, ²isa@uhamka.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 1 Juli 2023

Direvisi, 5 Juli 2023

Diiterima, 7 Juli 2023

Kata Kunci:

*Analisis Sentimen,
Algoritma Naïve Bayes,
Aplikasi TIX-ID,
Google Play Store,
Android*

ABSTRAK

Abstract- The TIX-ID application is an application-based online cinema ticket booking application. The TIX-ID application can be downloaded and accessed anywhere and anytime. The TIX-ID application is available on the Google Play Store and Appstore. How many users really like the TIX-ID application and however there are still some users who are dissatisfied with the services and features available in the TIX-ID application. So from these problems, sentiment analysis research was carried out to produce a classification of user satisfaction with the TIX-ID application using the Naïve Bayes algorithm method which has a fairly high accuracy value with a simple and fast process. The data obtained from data processing at the web scraping stage is as much as 1,600 data from user comments with a total of 1,259 positive sentiments and a total of 334 negative sentiments. The accuracy value obtained after implementing the Naïve Bayes algorithm obtained a percentage of 71.77%. The results of this percentage value indicate that to carry out a review sentiment analysis using the Naïve Bayes algorithm is very easy and suitable for use.

Abstrak- Aplikasi TIX-ID merupakan sebuah aplikasi pemesanan tiket bioskop *online* berbasis aplikasi. Aplikasi TIX-ID ini bisa di *download* dan di akses dimana saja dan kapan saja. Aplikasi TIX-ID ini tersedia di *Google Play Store* dan juga *Appstore*. Berapa para *user* sangat menyukai aplikasi TIX-ID tersebut dan akan tetapi masih ada beberapa para *user* yang tidak puas akan layanan dan fitur fitur yang telah tersedia dalam aplikasi TIX-ID. Maka dari permasalahan tersebut dilakukan penelitian analisis sentimen untuk menghasilkan sebuah klasifikasi terhadap kepuasan para *user* terhadap aplikasi TIX-ID dengan menggunakan metode algoritma *Naïve Bayes* yang memiliki nilai akurasi yang cukup tinggi dengan proses yang sederhana dan cepat. Data yang didapat dari pengolahan data pada tahap *web scraping* yaitu sebanyak 1.600 data ulasan komentar para *user* dengan total sentimen positif sebanyak 1.259 data dan total sentimen negatif sebanyak 334 data. Nilai akurasi yang didapat setelah melakukan implementasi algoritma *Naïve Bayes* diperoleh presentase sebesar 71,77%. Dengan hasil nilai presentase tersebut menandakan bahwa untuk melakukan sebuah analisis sentimen ulasan menggunakan algoritma *Naïve Bayes* sangat lah mudah dan cocok untuk digunakan.

Copyright © 2023 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Isa Faqihuddin Hanif

Program Studi Teknik Informatika,

Universitas Muhammadiyah Prof. Dr. Hamka

Jl. Tanah Merdeka No.6, Kampung Rambutan Ciracas Jakarta Timur, Indonesia

Email: isa@uhamka.ac.id

1. Pendahuluan

Kemampuan untuk mendapatkan informasi dengan cepat dan benar difasilitasi oleh perkembangan teknologi internet yang terus berkembang. Inovasi ini mendorong individu untuk merangkul dan memanfaatkan teknologi internet dalam kehidupan sehari-hari [1].

Teknologi berbasis *mobile* dalam operasional sehari-hari saat ini harus di implementasikan untuk mencapai efisiensi dari waktu ke waktu tanpa harus mengurangi komponen penting dari operasional itu sendiri [2].

Android adalah sistem operasi seluler yang berbasis *Linux* dan digunakan untuk perangkat genggam atau tablet (juga dikenal sebagai PDA). Fakta bahwa *Android* memberi pengembang akses ke platform terbuka yang dimana mereka dapat membangun aplikasi mereka sendiri setelah berkontribusi pada status platform tersebut sebagai sistem operasi *smartphone* paling populer pada saat ini. *Android* kini menjadi sistem operasi yang dominan untuk perangkat seluler. *Android* telah diberi kesempatan untuk melanjutkan ekspansinya yang pesat, dan kini menjadi salah satu sistem operasi untuk telepon seluler dan perangkat elektronik lainnya yang memiliki target pasar terbesar di dunia. *Android* adalah sistem operasi untuk perangkat seluler yang dibangun di atas platform *Linux*. Ini terdiri dari sistem operasi, *middleware*, dan aplikasi [2].

Google Play Store adalah platform khusus untuk distribusi digital yang dioperasikan oleh *Google*. *Google Play Store* adalah toko aplikasi resmi untuk sistem operasi *Android*. Pengguna dapat menjelajahi toko dan mengunduh aplikasi yang dikembangkan menggunakan *Android Software Development Kit (SDK)*, yang disediakan oleh *Google*. *Google Play Store* juga berfungsi sebagai toko media digital, memberi pelanggan akses ke buku, film, dan materi lainnya bersama dengan versi digital dari perangkat lunak populer seperti TIX-ID.

TIX-ID adalah sebuah layanan pemesanan tiket bioskop online berbasis aplikasi. Penggunaan aplikasi TIX-ID ini sangat digemari atau disenangi oleh para kalangan remaja dan orang dewasa, bukan hanya aplikasinya yang mudah digunakan tetapi aplikasi ini benar-benar sangat membantu para *user* jika ingin membeli sebuah tiket bioskop dan para *user* bisa mengaksesnya dengan sangat mudah dan tidak perlu mengantri lagi di loket penjualan tiket bioskop. Aplikasi ini sudah di *download* sekitar 10 juta pengguna dan sudah sekitar 455 ribu ulasan yang di berikan oleh para pengguna dalam memberikan ulasan dan kepuasan terhadap aplikasi TIX-ID.

Data utama dalam penelitian ini peneliti menggabungkan beberapa komentar dari para *user* terhadap aplikasi TIX-ID dalam *Google Play Store*

dengan menggunakan sebuah *tools* yaitu sebuah *Web Scraping* dengan memasukan beberapa kodingan dan data akan terkumpul dalam sebuah *file xls* yang kemudian data *file xls* tersebut peneliti gunakan untuk membuat penelitian ini.

Untuk memperbaiki dan meningkatkan akurasi temuan klasifikasi teks ulasan, Metode klasifikasi *Naïve Bayes* digunakan dalam penelitian ini untuk menilai penilaian pengguna terhadap aplikasi TIX-ID pada *website Google Play Store*. Informasi tersebut kemudian dapat digunakan sebagai acuan untuk mempertahankan popularitas dan kualitas layanan serta membangun kerangka kelemahan dan penilaian secara lebih positif sebagai hasil dari perolehan informasi lebih baik dan lebih akurat [3]. Penelitian terdahulu yang membahas tentang analisis sentimen menggunakan *Naïve Bayes* yaitu penelitian dari Siti Nurwahyuni dengan judul penelitian “Analisis Sentimen Aplikasi Transportasi Online KRL Access Menggunakan Metode *Naïve Bayes*” dengan temuan studi kasus tersebut maka hasil nilai akurasi untuk data uji ulasan pendapat yang menguntungkan dalam bahasa Indonesia untuk memilih aplikasi transportasi darat untuk *smartphone* adalah sebesar 84,00% ketika menggunakan algoritma NBC [4].

Berdasarkan penelitian yang dilakukan, peneliti mengharapkan hasil analisis sentimen ini dapat digunakan untuk melakukan evaluasi terhadap Aplikasi TIX-ID serta dapat mengetahui bagaimana ulasan atau tanggapan para *user* terhadap layanan dan fitur yang ada pada Aplikasi TIX-ID saat ini.

2. Metode Penelitian

Pada penelitian ini peneliti menggunakan metode algoritma *Naïve Bayes*. Didalam algoritma *Naïve Bayes* ini terdapat beberapa proses untuk melakukan klasifikasi terhadap ulasan komentar para *user* terhadap aplikasi TIX-ID yang akan di proses pada Rapid Miner.



Gambar 1. Alur Penelitian

2.1 Pengumpulan dan Pelabelan Dataset

Pada tahap ini peneliti melakukan pengumpulan dataset ulasan komentar para *user* terhadap aplikasi TIX-ID pada *Google Play Store* yang dimasukan kedalam *file excel* dan akan di proses pada Rapid Miner.



Gambar 2. Aplikasi Pengumpulan Dataset

Untuk tahap pelabelan dataset pada penelitian kali ini akan dilakukan secara manual yang akan dilakukan oleh 3 orang Mahasiswi dengan Program Studi Pendidikan Bahasa dan Sastra Indonesia dari Universitas Muhammadiyah Prof. Dr. Hamka dengan menentukan ulasan para *user* menjadi dua sentimen yaitu sentimen positif dan sentimen negatif. Peneliti menggunakan 1.600 data untuk melakukan pelabelan dataset pada ulasan komentar para *user* terhadap aplikasi TIX-ID lalu peneliti menggunakan Rapid Miner yang berfungsi untuk mengelompokkan sesuai kelas sentimen nya masing masing dengan ulasan positif dan ulasan negatif.

2.2 Preprocessing Dataset

Pada bagian *preprocessing* dataset berfungsi untuk mengubah data yang tidak terstruktur menjadi lebih terstruktur. Pada tahap ini terdapat beberapa macam yaitu *Tokenization*, *Transform Cases*, dan *Stopword Removal* [5].

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Tujuan dari *preprocessing* dataset adalah untuk menggabungkan beberapa bentuk kata yang berbeda menjadi satu bentuk kata. Prosedur yang dikenal sebagai *text preprocessing* melibatkan penempatan sebuah teks ke dalam satu atau lebih dari kategori yang telah ditentukan sebelumnya tergantung pada informasi di dalamnya. Selain itu, *preprocessing* teks mengacu pada klasifikasi secara otomatis materi digital yang ditemukan di internet, berita, artikel dan lain lain [6].

2.2.1 Tokenization

Tokenisasi data teks mengacu pada proses segmentasi kumpulan frasa menjadi karakter atau kata individual sesuai kebutuhan yang terkadang disebut sebagai token. Proses segmentasi ini mengubah kalimat menjadi kata-kata yang memiliki makna tertentu [7].

2.2.2 Transform Cases

Proses ini berfungsi untuk menormalkan teks menjadi satu huruf besar atau kecil [8].

2.2.3 Stopword Removal

Tujuan dari metode ini adalah untuk menghilangkan kata-kata yang tidak perlu dari teks. *Stopwords* adalah kata-kata yang tidak memiliki konteks dan dapat dihilangkan dengan menggunakan metode *bag-of-words* [9]. Jika ditemukan kalimat yang tidak masuk kedalam kategori *stopword* maka kalimat tersebut akan di hapus. Daftar *stopword* menggunakan *stopword list* sastrawi [10].

2.3 Algoritma Naïve Bayes

Pengklasifikasi *Naïve Bayes* adalah contoh langsung dari pengklasifikasi keluarga probabilitas. Itu didasarkan pada teorema Bayes dan beroperasi pada premis bahwa karakteristik individu sangat independen satu sama lain. Pengklasifikasi Naive Bayes sangat terukur karena hanya membutuhkan beberapa kecil parameter linier untuk memperhitungkan sejumlah besar prediktor atau fitur yang merupakan variabel dalam masalah pembelajaran. Ini adalah pengklasifikasi probabilitas yang paling sederhana dan tercepat terutama untuk fase pelatihan [11].

$$P(C_i|X) = \frac{p(X|C_i)p(C_i)}{p(X)} \quad (1)$$

Keterangan :

$P(C_i|X)$ = Probabilitas hipotesa C_i jika diberikan fakta atau *record X*

$P(X|C_i)$ = Mencari nilai parameter yang memberi kemungkinan yang paling besar

$P(X)$ = Jumlah *probability tuple* yang muncul [12].

2.4 Penerapan Algoritma

Dataset yang telah melewati tahap *preprocessing* akan diolah dengan algoritma *Naïve Bayes*. Pada tahapan ini, dilakukan perhitungan *Naïve Bayes* untuk penentuan ulasan para *user* dan mengelompokkan semua komentar yang mengandung ulasan positif dan ulasan negatif. Tahap awal cara kerja dari proses perhitungan *Naïve Bayes* adalah dengan melakukan pengambilan *data training* dari ulasan para *user* terhadap aplikasi TIX-ID.

2.5 Evaluasi

Tahap terakhir pada penelitian yaitu tahap evaluasi. Pada tahap ini merupakan tahap evaluasi terhadap hasil dari *preprocessing* dan implementasi algoritma yang digunakan oleh peneliti dengan berdasarkan *Accuracy*, *Precision*, dan *Recall*. Berdasarkan hasil perhitungan dari ketiga metode tersebut bahwa akan dilakukan analisis lebih lanjut dengan menggunakan algoritma *Naïve Bayes*. Analisis ini bertujuan untuk mengetahui apakah algoritma *Naïve Bayes* baik untuk digunakan dalam menganalisis sentimen terhadap ulasan para *user* terhadap aplikasi TIX-ID pada *Google Play Store*.

3. Hasil dan Pembahasan

3.1 Scraping Dataset

Pada proses ini peneliti melakukan *scraping* dataset yang bertujuan untuk menarik atau mengambil keseluruhan data yang akan di *labeling* dataset dan di proses pada Rapid Miner.

Web Scraping dapat mengubah data yang tidak terstruktur menjadi data terstruktur yang dapat disimpan dan divalidasi ke dalam database. Point dari web scraping adalah mengumpulkan data, menyimpan data, dan memvalidasi data. Langkah penting selanjutnya adalah analisis data, sehingga data yang telah divalidasi dapat diinterpretasikan menjadi informasi yang lebih baik [13].

```
#Scrape desired number of reviews
#Run kode ini jika ingin scrape data dengan jumlah tertentu. Ganti (misal, ingin

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'id.tix.android',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.MOST_RELEVANT, # defaults to Sort.MOST_RELEVANT you can use Sort.N
    count=100, # defaults to 100
    filter_score_with=None # defaults to None (means all score) Use 1 or 2 or 3 o
)
```

Gambar 3. Scraping Dataset

Setelah dilakukan *scraping* dataset dan *labeling* dataset maka hasil data positif yang dapat di ambil sebanyak 1.259 data dan data negatif yang dapat di ambil sebanyak 334 data.

3.2 Preprocessing Dataset

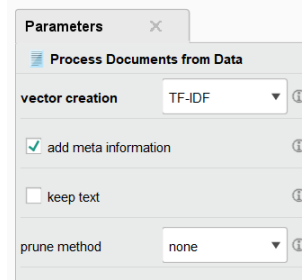
Pada proses ini peneliti melakukan *preprocessing* dataset yang membutuhkan proses *Tokenize*, *Transform Cases*, dan juga *Stopword Removal* yang berfungsi untuk mengolah dan menghilangkan data yang tidak termasuk ke dalam *class* positif dan *class* negatif.



Gambar 4. Preprocessing Dataset

3.3 Pembobotan Kata

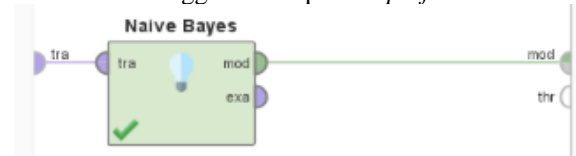
Pada fase ini kata-kata di kolom komentar aplikasi TIX-ID diberikan bobot yang kemudian melakukan tahap *preprocessing*. Tujuannya adalah untuk membuat *vector creation* TF-IDF menggunakan operator *process document* yang kita buat sebelumnya.



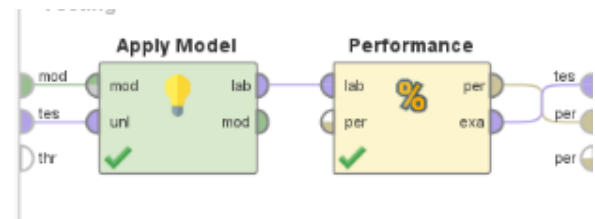
Gambar 5. Pembobotan Kata TF-IDF

3.4 Penerapan Algoritma Naïve Bayes

Kemudian pada proses ini peneliti melakukan penerapan algoritma *Naïve Bayes*. Data yang telah di *labeling* akan dihubungkan dengan algoritma *Naïve Bayes* untuk dilakukan sebuah evaluasi pada penerapan algoritma menggunakan validasi *Apply Model* dan menggunakan operator *performance*.



Gambar 6. Cross Validation



Gambar 7. Cross Validation

3.5 Evaluasi

Setelah semua operator dieksekusi, hasil *output* dari proses algoritma *Naïve Bayes* ini yang dihasilkan oleh aplikasi Rapid miner seperti gambar di bawah ini yang dimana hasil penerapan algoritma *Naïve Bayes* ini menunjukkan skor akurasi pada *cross validation* dengan hasil nilai akurasi sebesar 71.77%.

	True POSITIF	True Negatif	Class Precision
Pred. POSITIF	6111	65	90.38%
Pred. NEGATIF	276	256	48.12%
Class Recall	68.88%	79.75%	

Gambar 8. Hasil Penerapan Algoritma *Naïve Bayes*

Berdasarkan pengolahan data, mayoritas *user* pengguna TIX-ID puas dengan pelayanan serta fitur-fitur yang telah disediakan oleh para pengembang aplikasi TIX-ID, namun masih ada beberapa *user* pengguna yang tidak puas dengan pelayanan dan fitur yang telah tersedia di aplikasi TIX-ID.

4. Kesimpulan

Kesimpulan yang dapat di ambil dari penelitian kali ini yaitu bahwa penerapan algoritma *Naïve Bayes* sangat cocok untuk melakukan sebuah penilaian akurasi pada sebuah proses analisis sentimen ulasan

user terhadap aplikasi TIX-ID pada *Google Play Store*. Dengan hasil nilai akurasi sebesar 71.77% dengan data positif sebanyak 1.259 data dan data negatif sebanyak 334 data yang menandakan bahwa user pengguna aplikasi TIX-ID sangat puas akan dengan layanan dan fitur-fitur yang telah tersedia di dalam aplikasi TIX-ID. Kemudian untuk saran penelitian berikutnya mungkin para peneliti bisa menambahkan operator *optimize* pada model Rapid Miner agar hasil presentase nilai *accuracy* bisa menjadi lebih besar dan lebih baik.

5. Daftar Pustaka

- [1] I. R. Afandi, I. F. Hanif, F. N. Hasan, E. Sinduningrum, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression," vol. 5, no. 2, pp. 77–84, 2022.
- [2] I. F. Hanif, "Perancangan Aplikasi Android Penilaian Kinerja Dan Sikap Spiritual Karyawan (Studi Kasus Toko Retail Indomart)," *Pros. Semin. Nas. Teknoka*, vol. 4, no. 2502, pp. I91–I98, 2019, doi: 10.22236/teknoka.v4i0.4212.
- [3] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [4] S. Nurwahyuni, "Analisis Sentimen Aplikasi Transportasi Online Krl Access Menggunakan Metode Naive Bayes," *Swabumi*, vol. 7, no. 1, pp. 31–36, 2019, doi: 10.31294/swabumi.v7i1.5575.
- [5] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [6] A. I. Kadhim, "An Evaluation of Preprocessing Techniques for Text Classification," *Int. J. Comput. Sci. Inf. Secur.*, vol. 16, no. 6, pp. 22–32, 2018, [Online]. Available: <https://sites.google.com/site/ijcsis/>
- [7] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [8] V. Kalra and R. Aggarwal, "Importance of Text Data Preprocessing & Implementation in RapidMiner," *Proc. First Int. Conf. Inf. Technol. Knowl. Manag.*, vol. 14, no. January, pp. 71–75, 2018, doi: 10.15439/2017km46.
- [9] A. Riyani, M. Zidny, and A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," vol. 2, no. 1, pp. 23–27, 2019.
- [10] S. Khomsah and Agus Sasmito Aribowo, "Text-Preprocessing Model Youtube Comments in Indonesian," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 4, pp. 648–654, 2020, doi: 10.29207/resti.v4i4.2035.
- [11] U. N. Dulhare, "Prediction system for heart disease using Naive Bayes and particle swarm optimization," *Biomed. Res.*, vol. 29, no. 12, pp. 2646–2649, 2018, doi: 10.4066/biomedicalresearch.29-18-620.
- [12] D. Ayuningsih and N. A. Hasibuan, "Sistem Pakar Mendiagnosa Kerusakan Pada Mesin Penggilingan Padi Menggunakan Metode Naive Bayes," *J. JURIKOM (Jurnal Ris. Komputer)*, vol. 5, no. 4, pp. 371–376, 2018.
- [13] L. C. Dewi, Meiliana, and A. Chandra, "Social media web scraping using social media developers API and regex," *Procedia Comput. Sci.*, vol. 157, pp. 444–449, 2019, doi: 10.1016/j.procs.2019.08.237.