

Analisis Sentimen Pengguna Aplikasi Kitalulus pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes

Dwi Fristtikasari^{1*)}, Syariful Alam²⁾, Imay Kurniawan³⁾

¹⁾²⁾³⁾ Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana

^{*)}Correspondence author: dwifristtikasari39@wastukencana.ac.id, Purwakarta, Indonesia

DOI: <https://doi.org/10.37012/jtik.v10i2.2244>

Abstrak

Dalam beberapa tahun ini, masalah yang signifikan terkait dengan lowongan kerja yang sedang ramai terjadi di berbagai negara di seluruh dunia. Latar belakang utama dari fenomena ini tingginya tingkat pengangguran di beberapa wilayah, meskipun ada peningkatan ekonomi yang terjadi. Dalam era digital ini, ulasan pengguna di Google Play Store tidak hanya sekedar menyampaikan kesan individual, tetapi juga menjadi sumber informasi berharga bagi calon pengguna untuk memilih aplikasi yang sesuai dengan kebutuhan mereka. Ulasan pengguna merupakan salah satu bentuk timbal balik dari pengguna yang sepatutnya dianalisis oleh pihak pengembang agar dapat digunakan sebagai dasar pengembangan aplikasi. Analisis sentimen dapat dilakukan dengan pendekatan klasifikasi. Analisis sentimen dari ulasan pengguna dapat menjadi salah satu cara untuk mengetahui sentimen pengguna terhadap aplikasi. Penelitian ini bertujuan untuk menganalisis ulasan pada aplikasi kitalulus. Metode yang digunakan Naïve Bayes dengan Algoritma Multinomial Naïve Bayes. Dari hasil penelitian dapat disimpulkan terdapat analisis sentimen pengguna aplikasi KitaLulus ke dalam kategori positif dan negatif yang diambil melalui scraping dari bulan Oktober 2023 hingga Februari 2024 sebanyak 2.277 data ulasan dan diberi label positif dan negatif. Didapatkan prediksi klasifikasi sentimen pada data testing didapatkan hasil positif sebanyak 2094 dan sentimen negatif sebanyak 183 data ulasan, dan hasil klasifikasi dari metode naïve bayes didapat hasil accuracy 94%, precision 70%, dan tingkat keberhasilan (recall) 45%.

Kata Kunci: Analisis sentimen, KitaLulus, Google PlayStore, Naïve Bayes

Abstract

In recent years, significant problems related to busy job vacancies have occurred in various countries around the world. The main background to this phenomenon is the high level of unemployment in several regions, despite the economic improvement that has occurred. In this digital era, user reviews on the Google Play Store not only convey individual impressions, but also become a valuable source of information for potential users to choose applications that suit their needs. User reviews are a form of feedback from users which should be analyzed by the developer so that they can be used as a basis for application development. Sentiment analysis can be done using a classification approach. Sentiment analysis from user reviews can be one way to find out user sentiment towards an application. This research aims to analyze reviews on the Kitalulus application. The method used is Naïve Bayes with the Multinomial Naïve Bayes Algorithm. From the research results, it can be concluded that there is an analysis of the sentiment of users of the KitaLulus application into positive and negative categories taken through scraping from October 2023 to February 2024 of 2,277 review data and labeled positive and negative. The sentiment classification predictions obtained from the testing data obtained 2094 positive results and 183 negative sentiment results from the review data, and the classification results from the naïve Bayes method obtained an accuracy of 94%, a precision of 70%, and a success rate (recall) of 45%.

Keywords: Sentiment analysis, KitaLulus, Google PlayStore, Naïve Bayes

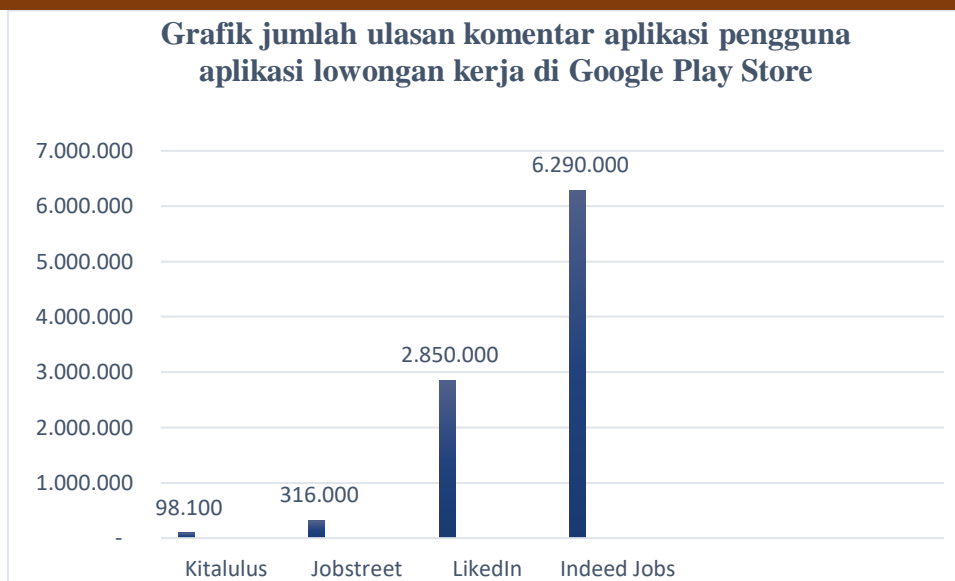
PENDAHULUAN

Dalam era digital ini, ulasan pengguna di Google Play Store tidak hanya sekadar menyampaikan kesan individual, tetapi juga menjadi sumber informasi berharga bagi calon pengguna untuk memilih aplikasi yang sesuai dengan kebutuhan mereka.

Dalam beberapa tahun ini, masalah yang signifikan terkait dengan lowongan kerja yang sedang ramai terjadi di berbagai negara di seluruh dunia. Latar belakang utama dari fenomena ini tingginya tingkat pengangguran di beberapa wilayah, meskipun ada peningkatan ekonomi yang terjadi. Tingkat pengangguran yang tinggi seringkali menjadi hasil dari ketidakcocokan antara keterampilan yang dimiliki oleh pencari kerja dengan permintaan yang ada di pasar tenaga kerja, yang sering kali berubah seiring dengan perkembangan teknologi dan ekonomi (Suarna & Prihartono, 2024).

Peningkatan persaingan di pasar kerja juga menjadi fenomena yang patut diperhatikan. Dengan jumlah pelamar yang lebih besar untuk setiap posisi, pelamar harus bersaing lebih keras untuk memperoleh pekerjaan yang diinginkan. Hal ini dapat mengakibatkan peningkatan tekanan dan ketidakpastian bagi para pencari kerja. Selain itu, meningkatnya penggunaan pekerja kontrak atau sementara oleh perusahaan juga menjadi masalah, karena hal ini seringkali berdampak negatif pada kestabilan dan keamanan pekerjaan (Rahayu & Indrati, 2024).

Aplikasi KitaLulus merupakan salah satu platform yang memberikan layanan dalam mencari lowongan kerja dan berbagai informasi terkait karier. Aplikasi ini sudah dirilis pada di tahun 2020, yaitu tepat saat masa pandemi. Misi utama KitaLulus, yaitu meningkatkan kesejahteraan masyarakat dengan memaksimalkan kesempatan kerja setiap individu lewat akses terhadap pendidikan berbasis komunitas dan teknologi (KitaLulus, 2024).



Gambar 1. Statistik Ulasan Pengguna Aplikasi Lowongan Kerja

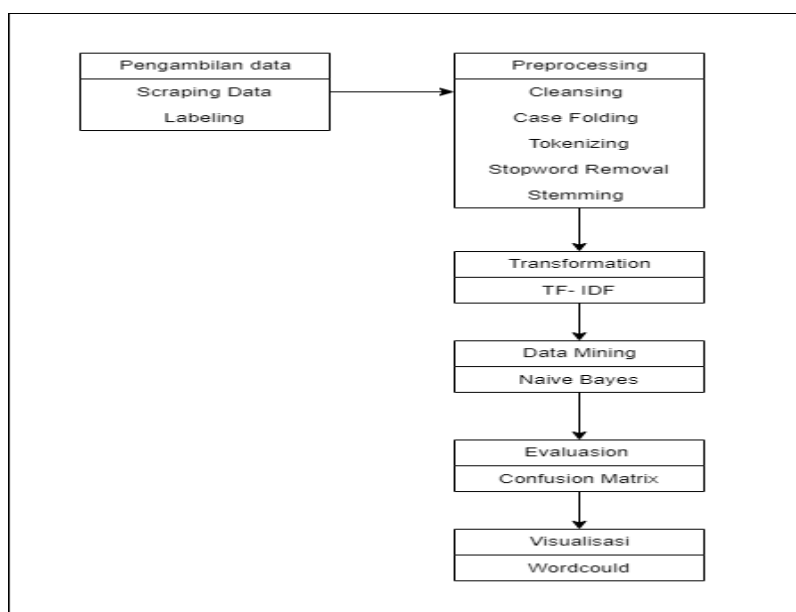
Berdasarkan observasi awal yang dilakukan terhadap aplikasi lowongan kerja yang terdapat di Google Play Store penulis melakukan riset beberapa aplikasi yang mana memiliki score di atas 4 diantaranya, yaitu: Jobstreet, Kitalulus, LinkedIn, Indeed Jobs. Dari keempat aplikasi lowongan kerja di google play store penulis akan meneliti Kitalulus. karena aplikasi kitalulus memiliki ulasan yang paling sedikit dan cenderung masih baru sehingga penelitian mengenai kitalulus belum banyak diteliti. Penelitian analisis sentimen pengguna aplikasi KitaLulus menggunakan algoritma Naïve Bayes penulis akan menganalisis dan menjabarkan bagaimana gambaran umum persepsi pengguna aplikasi Kitalulus berdasarkan ulasan terkait dampak yang diberikan oleh aplikasi tersebut baik positif maupun negatif. Analisis bertujuan untuk memberikan informasi kepada para pengguna aplikasi ke para pengguna dan pihak yang bersangkutan agar dapat mengelola dan menggunakan aplikasi dengan baik. Pada analisis sentimen penelitian dengan menerapkan algoritma Naïve Bayes dengan bantuan bahasa pemrograman python. Algoritma Naïve Bayes akan dimanfaatkan sebagai metode untuk menghasilkan prediksi terhadap suatu ulasan, apakah dalam kategori positif dan negatif.

Dalam Google Play Store terdapat informasi-informasi berupa deskripsi, komentar dari user dan rating mengenai aplikasi didalamnya dengan tujuan mengetahui kelebihan dan kekurangan dari aplikasi yang dibuat.

Evaluasi pengalaman pengguna menggunakan analisis adalah sentimen untuk menilai kepuasan dan ketidakpuasan pengguna KitaLulus melalui ulasan di Play Store. Ini memungkinkan mereka untuk menunjukkan area mana aplikasi dapat ditingkatkan atau diperbaiki. Meningkatkan pengembangan aplikasi meningkatkan pemahaman pengembang KitaLulus tentang kebutuhan dan preferensi pengguna untuk membantu pengembangan aplikasi dan peningkatan kualitasnya. Memberi Informasi untuk Keputusan Bisnis: Beri penyedia layanan Telegram dan pihak terkait informasi berharga untuk membantu membuat keputusan bisnis yang lebih baik, seperti strategi pemasaran, peningkatan fitur, dan upaya retensi pengguna (Suarna & Prihartono, 2024) .

METODE

Kerangka berpikir adalah logika penelitian yang diciptakan melalui pengamatan, fakta, dan tinjauan pustaka. Dalam penelitian ini, dijelaskan dalam kerangka berpikir sebagai berikut :



Gambar 2. Kerangka Berpikir Penelitian

1. Pengambilan Data

Pada penelitian ini, data yang didapatkan menggunakan teknik Web Scraping yang menggunakan pemrograman Python melalui sumber data ulasan pada aplikasi kitalulus

dalam Google Play Store. Data yang didapatkan sejumlah 2.388 yang meliputi ulasan dalam bentuk format CSV. Rentang waktu yang diambil adalah dari bulan Oktober 2023 sampai Februari 2024.

2. Labelling data

Data ulasan yang sudah di *scraping* akan diberi label positif dan negatif. Dari hasil pelabelan didapatkan data ulasan positif dan data ulasan negatif. Setelah itu, dilakukan pemberian nilai terhadap positif akan bernilai 1 dan jika negatif maka bernilai 0.

3. *Data Cleansing*

merupakan proses untuk menghilangkan atribut yang tidak berpengaruh terhadap klasifikasi yaitu tanda baca, karakter kosong, dan emoji

4. *Case Folding*

Case Folding adalah proses mengubah semua huruf data teks menjadi huruf kecil.

5. *Tokenization*

Proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian dianalisa. kata, angka, simbol, tanda baca, dan entitas penting lainnya dapat dianggap sebagai token.

6. *Stopword removal*

Merupakan proses untuk menghilangkan kata-kata yang dianggap tidak berpengaruh terhadap kalimat.

7. *Stemming*

Merupakan proses untuk mengubah kata-kata yang ada menjadi bentuk kata dasar.

8. *Term Frequency-Inverse Document Frequency (TF-IDF)*

TF-IDF yaitu pembobotan kata atau pemberian bobot suatu kata kedalam dokumen yang ada. TF (*Term Frequency*) adalah frekuensi kemunculan dari term yang ada pada dokumen, dan IDF (*Inverse Document Frequency*) yaitu sebuah perhitungan dari term untuk secara luas didistribusikan pada dokumen yang ada. TF (*Term Frequency*) akan terlebih dahulu dilakukan dalam menghitung proses pembobotan kata.

9. *Naïve Bayes*

Algoritma Naïve Bayes lebih memprioritaskan perhitungan probabilitas karena *Naïve Bayes* merupakan algoritma yang paling efektif untuk melakukan perhitungan dibandingkan algoritma yang lain.

10. *Confusion Matrix*

Confusion matrix bertujuan untuk mengetahui sejauh mana tingkat akurasi hasil klasifikasi data menggunakan metode *naïve bayes* dengan hasil evaluasi berupa *accuracy*, *precision*, dan *recall*.

11. *Visualization*

Tahap akhir yaitu tahap visualisasi, pada tahap ini peneliti menggunakan *wordcloud* untuk melihat kata apa saja yang sering muncul pada setiap statement yaitu statement negatif dan statement positif.

Analisis sentimen merupakan penelitian berupa text mining yang memiliki tujuan sebagai analisis pendapat, evaluasi, sikap, sentimen seseorang yang berupa teks untuk ditinjau pada suatu objek apakah kalimat tersebut termasuk kedalam kalimat positif atau kedalam kalimat negatif (Romadhoni et al., 2022).

Analisis sentimen adalah suatu proses yang bertujuan untuk mengetahui apakah polaritas data berupa teks (dokumen, kalimat, paragraf) akan mengarah ke positif, negative, atau netral. Analisis sentiment terdiri dari pemrosesan Bahasa alami, analisis teks dan komputasi linguistic untuk mengidentifikasi sentiment dari suatu dokumen informasi tekstual secara umum dapat dibagi menjadi informasi fakta dan opini. Tugas besar dalam analisis sentiment adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapatan yang dikemukakan dalam dokumen bersifat positif, negative atau netral. Penelitian mengenai analisis sentimen telah berkembang sejak tahun 2003 dan merupakan bagian dari text mining yang merupakan penelitian komputasi berdasarkan sentimen, emoticon, pendapat, komentar dan setiap ekspresi yang diungkapkan oleh teks (Fitriyana et al., 2023).

Naïve Bayes adalah metode yang tidak memiliki aturan dan menggunakan cabang matematika yang disebut teori probabilitas untuk mendapatkan peluang setinggi mungkin dengan melihat frekuensi atau jumlah kemunculan setiap klasifikasi dalam data pelatihan.

Dalam pengembangan basis data, Naïve Bayes melibatkan pembelajaran yang diawasi jenis pembelajaran mesin yang membutuhkan sampel sebagai data pelatihan label. dikelompokkan menjadi dua bagian, yaitu. Klasifikasi dan regresi. Klasifikasi saat variabel menjadi kategori, panas dan dingin, sakit atau tidak sakit dll. Variabel berupa nilai riil seperti bobot, nilai uang dll contoh lainnya yaitu Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN). Naïve Bayes juga merupakan metode yang menggunakan teknik probabilistik, dimana satu fitur dan fitur lainnya dalam data yang sama tidak saling berhubungan (Khoirul dkk., 2023). Berikut persamaan NBC :

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)}$$

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X

P(H) = Probabilitas hipotesis H

P(X|H) = Probabilitas hipotesis X berdasarkan kondisi H

P(H) = Probabilitas hipotesis H

Data *scraping* juga dikenal sebagai *web scraping*, adalah metode otomatis untuk mengambil informasi dari situs web. Teknik data Scraping umumnya digunakan untuk mengumpulkan sejumlah besar data dari berbagai sumber online seperti situs *e-commerce*, situs berita, dan platform media sosial. Dalam proses data scraping ini, para peneliti mengekstrak berbagai informasi relevan seperti ulasan dan peringkat, serta menyimpan data dalam format file CSV. Secara umum, data komentar yang dihasilkan dari proses scraping terdiri dari data asli yang ada pada elemen XML atau HTML di halaman-halaman website (Irawan & Bahtiar, 2023).

Pre-processing pada dasarnya dataset yang diperoleh dari proses text mining memiliki struktur yang sembarang dan tidak beraturan. Oleh karena itu, sebelum dataset

dimasukan kedalam model data terlebih dahulu melalui tahap *preprocessing* data, tahapan ini merupakan awal yang akan dilalui dalam memproses teks (Anton Permana et al., 2023).

Pada penelitian ini akan dilakukan tahapan *Preprocessing* dengan tahapan yaitu:

1. *Labelling* Tahap ini adalah pelabelan data, keputusan diambil untuk mengelompokkan data menjadi dua kategori: positif dan negatif. Atribut score akan digunakan sebagai dasar pelabelan sentimen untuk data latih. Kategori sentimen yang digunakan pada penelitian ini adalah positif dan negatif. Oleh karenanya, ulasan dengan rentang skor 4 dan 5 akan diberi label sentimen positif, rentang skor 1 dan 2 akan diberi label sentimen negatif, dan data ulasan dengan rentang skor 3 akan dihapus karena dianggap sebagai sentimen netral dan tidak akan digunakan dalam penelitian ini (Arifiyanti et al., 2023)
2. *Data cleansing* Tahap ini adalah tahap eliminasi aksara non alfabetis untuk menurunkan noise. Aksara yang dihapus adalah tanda baca seperti titik (.), koma (,), tanda tanya (?), dan tanda seru (!), serta simbol-simbol seperti tanda '@' untuk username, hashtag (#), emoticon, dan alamat website.
3. *Case folding* adalah tahapan untuk menyeragamkan seluruh teks yang akan dimasukan kedalam model menjadi huruf kecil semua (lowercase).
4. *Tokenization* yaitu proses pemecahan kata pada kalimat review, pemisahan kata dalam kalimat umumnya menggunakan karakter spasi, maka karakter spasi diandalkan untuk proses tokenisasi ini.
5. *Stopword removal* adalah tahapan untuk menghilangkan kata yang tidak di perlukan
6. *Stemming* yaitu proses menemukan kata dasar dengan menghilangkan semua imbuhan yang menempel pada kata.

Word Cloud adalah gambaran visual berdasarkan frekuensi kemunculan kata-kata pada suatu kumpulan teks, dimana ukuran huruf menentukan frekuensi kemunculan sebuah kata yang artinya semakin besar ukuran huruf maka semakin besar kemunculan kata tersebut dan sebaliknya, semakin kecil huruf maka semakin kecil frekuensi kemunculan kata tersebut (Tri Julianto, 2022).



Gambar 3. Wordcloud (Wordcloudapi)

HASIL DAN PEMBAHASAN

Sebelum klasifikasi dataset komentar, gunakan sklearn library dan algoritma *multinomial Naive Bayes* untuk mengklasifikasikan komentar sehingga menjadi data clean, dan membaginya menjadi dua bagian, yaitu 80% dari dataset akan digunakan untuk proses pelatihan data, dan 20% sisanya akan digunakan sebagai bahan uji data.

Tabel 1. Data Sampel

| Teks ulasan | sentimen |
|--|----------|
| informasi lowongan kerja | Positif |
| aplikasinya bagus membantu mencari pekerjaan | Positif |
| sangat membantu dalam mencari pekerjaan semoga semakin baik lagi aplikasinya | Positif |
| apk keren berguna membatu mencari kerja | Positif |
| aplikasi bagus mencari informasi seputar lowongan pekerjaan | Positif |
| aplikasi penipuan percaya | Negatif |
| lokernya dikit banget gak update kalah akun ig ngeshare lowongan cepet diterima kalo cuman gantungan harapan orang aja | Negatif |
| kebanyakan sales | Negatif |
| aplikasi penipuan | Negatif |
| lowongan palsu alias penipuan kayak bri live | Negatif |

Sampel kumpulan data yang akan digunakan yaitu 10 ulasan, termasuk 10 sampel yang diberi label 5 positif dan 5 negatif. Data tersebut sebagai data latih. Seperti yang ditunjukkan pada tabel 1.

Tabel 2. Term Frequency

| No | Kata | Dokumen | | | | | | | | | |
|----|-------------|---------|----|----|----|----|----|----|----|----|-----|
| | | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 |
| 1 | informasi | 1 | | | | 1 | | | | | |
| 2 | lowongan | 1 | | | | 1 | | | | | 1 |
| 3 | kerja | 1 | | | 1 | | | | | | |
| 4 | aplikasinya | | 1 | | 1 | | | | | | |
| 5 | bagus | | 1 | | | 1 | | | | | |
| 6 | membantu | | 1 | 1 | 1 | | | | | | |
| 7 | mencari | | 1 | 1 | 1 | 1 | | | | | |
| 8 | pekerjaan | | 1 | 1 | | 1 | | | | | |
| 9 | sangat | | | 1 | | | | | | | |
| 10 | dalam | | | 1 | | | | | | | |
| 11 | semoga | | | 1 | | | | | 1 | | |
| 12 | semakin | | | 1 | | | | | | | |
| 13 | Baik | | | 1 | | 1 | | | | 1 | |
| 14 | lagi | | | 1 | | | | | | | |
| 15 | apk | | | | 1 | | | | | | |
| 16 | keren | | | | 1 | | | | | | |
| 17 | berguna | | | | 1 | | | | | | |
| 18 | aplikasi | | | | | 1 | 1 | | | 1 | |
| 19 | seputar | | | | | 1 | | | | | |
| 20 | penipuan | | | | | | 1 | | | 1 | 1 |
| 21 | percaya | | | | | | 1 | | | | |
| 22 | lokernya | | | | | | | 1 | | | |
| 23 | dikit | | | | | | | 1 | | | |
| 24 | banget | | | | | | | 1 | | | |
| 25 | gak | | | | | | | 1 | | | |
| 26 | update | | | | | | | 1 | | | |
| 27 | kalah | | | | | | | 1 | | | |
| 28 | akun | | | | | | | 1 | | | |
| 29 | ig | | | | | | | 1 | | | |
| 30 | ngeshare | | | | | | | 1 | | | |
| 31 | cepat | | | | | | | 1 | | | |
| 32 | diterima | | | | | | | 1 | | | |
| 33 | kalo | | | | | | | 1 | | | |
| 34 | cuma | | | | | | | 1 | | | |
| 35 | gantungan | | | | | | | 1 | | | |
| 36 | harapan | | | | | | | 1 | | | |
| 37 | orang | | | | | | | 1 | | | |
| 38 | aja | | | | | | | 1 | | | |
| 39 | kebanyakan | | | | | | | 1 | | | |
| 40 | sales | | | | | | | 1 | | | |
| 41 | palsu | | | | | | | | | | 1 |
| 42 | alias | | | | | | | | | | 1 |

| No | Kata | Dokumen | | | | | | | | | |
|----|-------|---------|----|----|----|----|----|----|----|----|-----|
| | | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 |
| 43 | kayak | | | | | | | | | | 1 |
| 44 | bri | | | | | | | | | | 1 |
| 45 | live | | | | | | | | | | 1 |

Untuk melakukan perhitungan menggunakan contoh dataset data tabel 1 menggunakan *Microsoft Excel*, langkah pertama adalah memastikan *term frequency* (TF) di setiap dokumen, yang ditunjukkan pada tabel 2.

Dari tabel 2 Selanjutnya dilakukan perhitungan dari kemunculan kata unik dan jumlah kata pada setiap kategori. Alur perhitungan dapat dilihat pada tabel 3:

Tabel 3. Jumlah Kata Unik

| Kategori | Jumlah dokumen di setiap kategori | Jumlah dokumen | Probabilitas (kategori) | Jumlah kata setiap kategori |
|------------------|-----------------------------------|----------------|-------------------------|-----------------------------|
| Positif | 5 | 10 | 5 | 31 |
| Negatif | 5 | 10 | 5 | 32 |
| Jumlah kata unik | | 45 | | |

Kata unik merupakan kata yang sering muncul setiap dokumen. Setelah mendapat jumlah kata unik dan jumlah kata pada setiap kategori, kemudian dilakukan perhitungan probabilitas pada setiap kategori.

$$P(\text{kata} | \text{kategori}) = \frac{\text{jumlah dari satu kata di satu kategori} + 1}{\text{jumlah kata tiap kategori} + \text{jumlah keseluruhan kata unik}}$$

Dari tabel 3 di atas dihitung jumlah kata, dalam kategori positif 31 kata, negatif 32 kata dengan jumlah kata unik sebanyak 45 kata. Untuk menghitung probabilitas pada setiap kategori adalah sebagai berikut :

$$(\text{pekerjaan} | \text{positif}) = \frac{3 + 1}{31 + 45} = \frac{4}{76} = 0,0526$$

$$(\text{penipuan} | \text{negatif}) = \frac{3 + 1}{32 + 45} = \frac{4}{77} = 0,0519$$

Pada tahapan ini, dilakukan perhitungan pada dataset, seperti pada tabel 4 di bawah ini:

Tabel 4. Hasil Perhitungan *Naive Bayes*

| Kata Unik | Positif | Negatif |
|------------------|----------------|----------------|
| informasi | 0,03947368 | - |
| lowongan | 0,03947368 | 0,02597403 |
| kerja | 0,03947368 | - |
| aplikasinya | 0,03947368 | - |
| bagus | 0,03947368 | - |
| membantu | 0,05263158 | - |
| mencari | 0,06578947 | - |
| pekerjaan | 0,05263158 | - |
| sangat | 0,02631579 | - |
| dalam | 0,02631579 | - |
| semoga | 0,02631579 | 0,02597403 |
| semakin | 0,02631579 | - |
| Baik | 0,03947368 | 0,02597403 |
| lagi | 0,02631579 | - |
| apk | 0,02631579 | - |
| keren | 0,02631579 | - |
| berguna | 0,02631579 | - |
| aplikasi | 0,02631579 | 0,03896104 |
| seputar | 0,02631579 | - |
| penipuan | - | 0,05194805 |
| percaya | - | 0,02597403 |
| lokernya | - | 0,02597403 |
| dikit | - | 0,02597403 |
| banget | - | 0,02597403 |
| gak | - | 0,02597403 |
| update | - | 0,02597403 |
| kalah | - | 0,02597403 |
| akun | - | 0,02597403 |
| ig | - | 0,02597403 |
| ngeshare | - | 0,02597403 |
| cepat | - | 0,02597403 |
| diterima | - | 0,02597403 |
| kalo | - | 0,02597403 |
| cuma | - | 0,02597403 |

| Kata Unik | Positif | Negatif |
|------------|---------|------------|
| gantungan | - | 0,02597403 |
| harapan | - | 0,02597403 |
| orang | - | 0,02597403 |
| aja | - | 0,02597403 |
| kebanyakan | - | 0,02597403 |
| sales | - | 0,02597403 |
| palsu | - | 0,02597403 |
| alias | - | 0,02597403 |
| kayak | - | 0,02597403 |
| bri | - | 0,02597403 |
| live | - | 0,02597403 |

Berikut adalah tahapan pengujian *Naive bayes* menggunakan python yang ditunjukkan pada gambar 2 di bawah ini :

```
MultinomialNB Accuracy: 0.9407894736842105
MultinomialNB Precision: 0.7083333333333334
MultinomialNB Recall: 0.4594594594594595
MultinomialNB f1_score: 0.5573770491803279
confusion_matrix:
[[ 17 20]
 [ 7 412]]
=====
              precision    recall  f1-score   support

   Negatif      0.71      0.46      0.56         37
   Positif      0.95      0.98      0.97        419

 accuracy              0.94         456
 macro avg      0.83      0.72      0.76         456
 weighted avg   0.93      0.94      0.93         456
```

Gambar 4. Alur Klasifikasi Semua Sentimen

Evaluasi merupakan bertujuan untuk mengecek kebenaran hasil klasifikasi. Hasil evaluasi menggunakan metode cross validation yang nantinya akan diukur dengan menggunakan confusion matrix untuk mengetahui hasil accuracy, precision, dan recall.

Berikut ini merupakan tabel dari confusion matrix hasil yang di dapatkan setelah dilakukan pengujian dari keeluruhan data ulasan dengan menggunakan Google Colabory dapat dilihat pada tabel dibawah ini.

Tabel 5. Confusion Matrix

| | True | False |
|---------|------|-------|
| Positif | 17 | 20 |
| Negatif | 7 | 412 |

Perhitungan Akurasi

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} = \frac{17 + 412}{(456)} = 0,94$$

Perhitungan Presisi

$$Precision = \frac{TP}{TP + FP} = \frac{17}{17 + 7} = 0,70$$

Perhitungan Recall

$$Recall = \frac{TP}{TP + FN} = \frac{17}{17 + 20} = 0,45$$

Visualisasi adalah tampilan berupa grafis atau visual dari informasi dan data. Mengubah data menjadi lebih sederhana untuk ditampilkan. Visualisasi digambarkan dengan *Word Cloud* pada keseluruhan sentimen positif dan negatif tentang ulasan aplikasi KitaLulus di Google play store. Hasil dari Visualisasi ditampilkan pada gambar dibawah ini:



Gambar 5. Word Cloud Sentimen Kitalulus

Pada gambar 5 *word cloud* diatas terdapat kata-kata yang sering muncul pada data ulasan KitaLulus keseluruhan adalah kata "aplikasi", "kita", "lulus" dan lain-lain.

tergolong positif dengan hasil 94% pada akurasi, 70% pada nilai precision, dan tingkat keberhasilan (recall) 45%. Sedangkan dari hasil visualisasi data ulasan negatif terdapat kata-kata "aplikasi", "kerja", "loker", "susah" dan lain-lain. Dapat disimpulkan bahwa aplikasi KitaLulus sering terjadi update. KitaLulus juga perlu meningkatkan kemampuan aplikasi dalam melakukan update tentang loker yang tidak lengkap kebanyakan loker sales dan banyak iklan.

Untuk pengembangan lebih lanjut dari penelitian ini, maka dapat disarankan sebagai berikut:

1. Untuk Peneliti selanjutnya bisa mencoba atau menggunakan algoritma yang lainnya dalam menganalisis sentimen seperti algoritma K-Nearest Neighbor, Random Forest, ataupun algoritma lainnya.
2. Dapat memperluas dataset yang lebih besar dan lebih beragam agar dapat membantu meningkatkan kinerja model analisis sentiment. Melibatkan data dari berbagai sumber, atau Bahasa agar dapat memberikan wawasan yang lebih luas tentang sentimen.

REFERENSI

- Anton Permana, M., Widiastuti, S., & Saepudin, S. (2023). *Analisis Sentimen Penggunaan Aplikasi Video konferensi Pada Ulastoko Bermain Google menggunakan Metode Nsm (Klasifikasi Naive Bayesugd)*.
- Arifiyanti, A., Shantika, R., & Syafira, O. (2023). *Analisis Sentimen Ulasan Pengguna Bsi Mobile Pada Google Play Dengan Pendekatan Supervised Learning*.
- Doloksaribu, HP, & Samuel, YT (2022). Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi. *Jurnal Teknologi Informasi: Jurnal ...*, e-journal.upr.ac.id, <https://e-journal.upr.ac.id/index.php/JTI/article/view/3747>
- Fitriyana, V., Hakim, L., Candra Rini Novitasari, D., Hanif Asyhar, A., (2023). Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine. In *Jurnal Buana Informatika* (Vol. 14, Issue 1).
- Hendra, A, & Fitriyani, F (2021). Analisis Sentimen Review Halodoc Menggunakan Nai've Bayes Classifier. *JISKA (Jurnal Informatika Sunan ...*, ejournal.uin-suka.ac.id, <https://ejournal.uin-suka.ac.id/saintek/JISKA/article/view/2076>

- Irawan, B., & Bahtiar, A. (2023). Penggunaan Algoritma Naïve Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi Adakami Di Google Play Store. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).
- Khoirul, M., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1).
- KitaLulus. (2024). *Platform Pencarian Kerja Berorientasi Komunitas Terbesar dan Teraman di Indonesia*. 2023. <https://www.kitalulus.com/tentang-kitalulus>
- Negara, ABP, Muhandi, H, & Putri, IM (2020). Analisis sentimen maskapai penerbangan menggunakan metode naive bayes dan seleksi fitur information gain. *J. Teknol. Inf. dan Ilmu Komput*, academia.edu, <https://www.academia.edu/download/68610303/pdf.pdf>
- Rahayu, G. Y., & Indrati, A. (2024). Perancangan Ulang Antarmuka Portal Kitalulus Dengan Menggunakan Metode User Centered Design (UCD). In *Januari* (Vol. 3, Issue 1).
- Romadhoni, Y., Fahmi, K., & Holle, H. (2022). *Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM*. 7(2).
- Suarna, N., & Prihartono, W. (2024). Penerapan NLP (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2).
- Salim, SS, & Mayary, J (2020). Analisis Sentimen pengguna Twitter terhadap dompet elektronik dengan metode lexicon based dan k-nearest neighbor. *Jurnal Ilmiah Informatika Komputer*, ejournal.gunadarma.ac.id, <https://ejournal.gunadarma.ac.id/index.php/infokom/article/view/2411>
- Tri Julianto, I. (2022). *Analisis Sentimen Terhadap Sistem Informasi Akademik Mahasiswa Institut Teknologi Garut*. <https://jurnal.itg.ac.id/>
- Tuhuteru, H (2020). Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma Support Vector Machine. *Journal Information System Development ...*, ejournal-medan.uph.edu, <https://ejournal-medan.uph.edu/isd/article/view/381>