(JUSS) Jurnal Sains dan Sistem Informasi ISSN 2614-8277
Vol. 4 No.1, Juli 2021

Analisis Sentimen Aplikasi Quara Menggunakan Algoritma Naive Bayes

Natasya Yonike Ambarita
[1]Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi
email: penulis [1]natasyayonikeambarita@gmail.com

Abstract

This project seeks to utilize the Naive Bayes algorithm to analyze the sentiment in user evaluations of the Quora app on the Google Play Store. This research offers further depth into users' perceptions of Quora using review sentiment analysis, which aids developers in enhancing the user experience. Data from user reviews on the Google Play Store are gathered for this study using the web scraping technique. Data collection, preprocessing, labeling, TF-IDF weighting, data split into training and test data, classification using the Nave Bayes method, and model assessment were the steps of the research. The findings demonstrate that the Naive Bayes algorithm is 88% accurate in classifying the sentiments of user reviews. It is advised to conduct further study to expand data volume, employ feature selection, and boost accuracy.

Keywords: *Quora, Sentiment Analysis, Naïve Bayes, TF-ID, Play Store*

1. PENDAHULUAN

Dalam era digital yang terus berkembang, aplikasi mobile telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari kita. Salah satu aplikasi yang telah memikat perhatian pengguna adalah Quora, sebuah platform tanya jawab yang memberdayakan orang-orang untuk berbagi dan mengembangkan pengetahuan dunia. Orang-orang datang ke Quora untuk mengajukan pertanyaan tentang berbagai subjek, membaca pengetahuan berkualitas tinggi yang dipersonalisasikan dan relevan bagi mereka, serta membagikan pengetahuan mereka dengan yang lain. Quora adalah tempat untuk berbagi pengetahuan dan memahami dunia lebih baik [1]. Dengan menyediakan akses ke berbagai topik dan pengetahuan dari berbagai sumber, Quora berhasil membangun popularitasnya di antara pengguna yang haus akan informasi.

Dalam ekosistem aplikasi mobile yang semakin kompleks, ulasan pengguna memainkan peran penting dalam membantu calon pengguna untuk memahami kualitas dan pengalaman pengguna lain dengan aplikasi tertentu. Ulasan pengguna di Google Play Store, sebagai salah satu platform utama untuk distribusi aplikasi Android, memberikan wawasan berharga tentang pengalaman pengguna dengan Quora.

Namun, dengan jumlah ulasan Quora yang terus meningkat di Google Play Store, menganalisis dan memahami sentimen yang terkandung dalam setiap ulasan secara manual menjadi tugas yang sangat sulit dan memakan waktu. Oleh karena itu, diperlukan pendekatan analisis sentimen yang efisien untuk memproses dan memahami ulasan-ulasan ini secara otomatis.

Analisis sentiment digunakan untuk mengetahui suatu pendapat atau tanggapan terkait masalah, mempunyai nilai pro atau kontra dan kemudian dijadikan sebagai referensi untuk memperbaiki layanan, atau memperbaiki penilaian produk [2]. Pendapat atau opini memiliki karateristik yang menunjukan bahwa hal tersebut bersifat baik atau buruk, peninjauan berbagai macam pendapat masyarakat diperlukan dikarenakan itu akan menjadi perwakilan dari masyarakat.

Dalam konteks ini, penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam analisis sentimen ulasan pengguna aplikasi Quora di Google Play Store. Analisis sentimen atau opinion mining bertujuan untuk mengetahui apakah opini yang terkandung dalam suatu kalimat adalah positif atau negatif dan Algoritma Naive Bayes Classifier dianggap cocok untuk digunakan dalam analisis sentimen karena bertujuan sebagai metode klasifikasikedalam

kategori positif dan negatif [3]. Dengan menerapkan algoritma ini, penelitian ini bertujuan untuk mendapatkan pemahaman yang lebih mendalam tentang pandangan pengguna terhadap aplikasi Quora.

Melalui analisis sentimen ulasan aplikasi Quora, penelitian ini memiliki beberapa manfaat penting. Pertama, penelitian ini akan memberikan wawasan yang lebih mendalam tentang tanggapan pengguna secara keseluruhan terhadap Quora. Hasil analisis ini dapat membantu pengembang dalam mengidentifikasi aspek-aspek yang dianggap positif atau negatif oleh pengguna, sehingga mereka dapat membuat keputusan yang tepat untuk meningkatkan pengalaman pengguna.

Selain itu, penelitian ini juga akan memberikan manfaat bagi calon pengguna dalam memahami pandangan pengguna lain terhadap Quora. Melalui analisis sentimen, calon pengguna dapat memperoleh informasi yang lebih objektif dan dapat dipercaya mengenai kelebihan dan kekurangan aplikasi ini, yang dapat membantu mereka dalam membuat keputusan yang lebih terinformasi dalam menggunakan atau mengunduh Quora.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi penting dalam pemahaman tentang pandangan dan sentimen pengguna terhadap aplikasi Quora di Google Play Store. Hasil penelitian ini dapat dimanfaatkan oleh pengembang aplikasi untuk meningkatkan kualitas dan pengalaman pengguna Quora, serta memberikan panduan bagi calon pengguna dalam memilih aplikasi yang sesuai dengan kebutuhan mereka.

2. METODE PENELITIAN

Penelitian ini menggunakan bantuan aplikasi Anaconda dengan vahasa pemrograman Python versi 3.9.12 dan menggunakan IDE Jupyter Notebook 6.0.3. Tahapan yang harus ditempuh pada penelitian ini berjumlah 7 tahap, yaitu pengumpulan data (*scarping* data), data *preprocessing,* pelabelan data, pembobotan TF-IDF, pembagian data dengan tujuan untuk membuat data latih dan data uji, klasifikasi menggunakan algoritma Naïve Bayes, dan evaluasi model.

2.1. Pengumpulan Data

Pada tahap pengumpulan data dengan menggunakan *web scraping*. Cara kerja *web scraping* adalah dengan mengakses halaman web, memilih elemen data yang ada dalam halaman tersebut, melakukan ekstraksi dan transformasi bila diperlukan, dan terakhir menyimpan data tersebut menjadi dataset terstruktur (Boeing and Waddell,2017). Data yang diambil pada proses *web scraping* merupakan data yang berasal dari halaman website dengan bantuan library Google-Play-Scraper menggunakan Google Colaboratory IDE. Jumlah data yang diambil sebanyak 1000 data dilakukan pada periode sebelum tanggal 08 Juni 2023.

2.2. Data Preprocessing

Dalam melakukan analisis sentimen, sebelumnya perlu melalui tahap data *preprocessing*. Tahap ini diperlukan untuk mengoptimalkan hasil dari analisis sentimen. Data *preprocessing* yang akan dilakukan terdapat lima tahap, yaitu *case folding,* remove punctuation, tokenizing, stopwords removal, dan *lemmatization.*

1. *Case folding* merupakan tahap pengolahan data untuk membuat semua huruf menjadi huruf kecil.
2. *Remove punctuation*
   Remove punctuation, atau penghapusan tanda baca. Pada langkah ini hanya huruf alfabet yang diterima sedangkan karakter selain huruf akan dihilangkan.
3. *Tokenizing*
   Tokenizing adalah proses dimana memecahkan kalimat menjadi kata – perkata atau disebut dengan proses pemenggalan kata. *Tokenizing* adalah tahap pemotong *string input* berdasarkan kata – kata yang tersusun. Token yang dihasilkan berupa kata tunggal yang akan memudahkan dalam proses selanjunya [4].
4. Stopwords removal
   Stopword adalah proses menghilangkan atau menghapus kata – kata hasil *tokenizing* yang tidak memiliki makna seperti “ke”,”di”, “mau”,”itu” dan sebagainya. *Stopword* adalah daftar kata – kata yang tidak menampilkan isi dari suatu dokumen teks *stopword* dilakukan untuk

 ISSN 2614-8277

menghapus kata atau *term* yang tidak memilki arti [5].

5. *Lemmazation*
   Lemmatization adalah proses dalam pemrosesan teks yang mengubah kata-kata dalam bentuk dasar.

2.3. Pelabelan data

Pelabelan data adalah proses memberikan label atau tag pada setiap data dalam dataset. Dalam konteks analisis sentimen, pelabelan data melibatkan memberikan label sentimen tertentu (misalnya positif, negatif, atau netral) pada setiap data teks dalam dataset. Pelabelan data dapat dilakukan dengan menggunakan bantuan skor yang terdapat pada ulasan. Dengan memberikan nilai positif, netral dan negatif terhadap skor 1 hingga 5.

2.4. Pembobotan TF-IDF

Proses selanjutnya, dilakukan pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), pembobotan kata berfungsi untuk mengubah data berupa teks menjadi numerik. Bidang pembelajaran mesin atau *Machine learning* dan *deep learning* akan bekerja dengan baik dan maksimal jika data yang diolah berupa data numerik. Maka setiap penelitian apda bidang ini khususnya analisis sentimen harus mengubah data berupa kumpulan kata-kata atau kalimat menjadi numerik. Proses tersebut bisasnya disebut dengan proses pembobotan kata. Proses ini memiliki banyak metode yang dapat digunakan, diantaranya BoF (Bag of Words, N-gram, Word2VecdanTF-IDF [6]. TF-IDF merupakan salah satu teknik yang digunakan dalam pengolahan teks untuk memberikan bobot pada kata-kata dalam sebuah dokumen. Tujuan dari TF-IDF adalah untuk mengidentifikasi kata-kata yang paling penting dalam suatu dokumen atau kumpulan dokumen. *Term Frequency* (TF) adalah nilai frekuensi kemunculan suatu kata dalam sebuah dokumen. Nilai TF dapat dihitung dengan menggunakan rumus :

$$TF = \frac{(jumlah\ kemunculan\ kata\ dalam\ dokumen)}{(jumlah\ kata\ dalam\ dokumen)}$$

Namun, nilai TF ini tidak memberikan informasi tentang pentingnya kata tersebut dalam dokumen. Kata - kata yang sering muncul dalam sebuah dokumen seperti kata hubung atau kata umum mungkin memiliki nilai TF yang tinggi, tetapi sebenarnya tidak memiliki makna yang penting dalam dokumen tersebut.Oleh karena itu, dibutuhkan teknik lain, yaitu *Inverse Document Frequency* (IDF) yang memberikan bobot pada kata-kata yang muncul jarang di seluruh dokumen. Nilai IDF dapat dihitung dengan menggunakan rumus:

$$\text{IDF} = log\ \frac{N}{n}$$

Dimana N adalah jumlah dokumen dalam kumpulan dokumen dan n adalah jumlah dokumen yang mengandung kata tersebut. Setelah nilai TF dan IDF diperoleh, nilai TF-IDF dapat dihitung dengan mengalikan nilai TF dengan nilai IDF. Kata-kata yang memiliki nilai TF-IDF yang tinggi dianggap penting dan memberikan kontribusi yang lebih besar dalam menentukan topik dokumen atau kumpulan dokumen tersebut [6].

2.5. Pembagian Data

Pembuatan data latih dan data uji bertujuan untuk membagi data dari hasil data preprocessing ke dalam perbandingan tertentu. Perbandingan jumlah pembagian data dalam penelitian ini adalah 8:2 yang terdiri dari 80% data latih atau sebanyak 800 data dan 20% data uji atau sebanyak 200 data. Pembagian dengan perbandingan tersebut dikarenakan semakin banyak data yang dilatih, maka akan menghasilkan hasil analisis yang lebih akurat [7].

2.6. Klasifikasi Menggunakan Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi yang didasarkan oleh teorema Bayes [8]. Naïve Bayes juga merupakan salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistika [9] Pada penelitian yang dilakukan menggunakan algoritma multinominal naïve bayes (MNB). Algoritma ini mengimplementasikan algoritma naïve bayes

untuk data yang didistribusikan secara multinasional dan merupakan salah satu dari dua varian naive bayes klasik yang digunakan dalam klasifikasi teks [10]. MNB cocok untuk mengklasifikasikan teks yang memiliki banyak fitur (misalnya kata-kata dalam dokumen) dengan menggunakan metode probabilitas. Algoritma ini bekerja dengan menghitung probabilitas kemunculan setiap kata di setiap kelas dan kemudian menggunakan probabilitas ini untuk memprediksi kelas baru dari suatu teks. Implementasi Multinominal Naïve Bayes biasayanya digunakan pada klasifikasi email sebagai spam atau bukan spam, klasifikasi dokumen sebagai topik tertentu, atau klasifikasi sentimen pada teks (positif, negatif, atau netral). Kode yang digunakan saat proses eksperimen dengan mengimport kelas MultinomialNB dari modul naïve bayes pada library scikit-learn. Dengan menggunakan MultinomialNB dapat melakukan pelatihan dan prediksi dengan mudah pada data teks.

Rumus Multinomial Naive Bayes digunakan untuk menghitung probabilitas dokumen terhadap setiap kategori yang ada. Probabilitas ini digunakan untuk memprediksi kategori mana yang paling mungkin untuk dokumen tersebut.

Rumus dasar Multinomial Naive Bayes adalah sebagai berikut :

$$P\left(\frac{c}{d}\right) = P(c) * P\left(\frac{d}{c}\right)/P(d)$$

Keterangan :

P = $\left(\frac{c}{d}\right)$ adalah probabilitas dokumen d masuk ke dalam kategori c.

P = (c) adalah probabilitas prior untuk kategori c.

P = $\left(\frac{d}{c}\right)$ adalah probabilitas kemunculan fitur (kata atau istilah) dalam dokumen d, jika dokumen d masuk dalam kategori c.

P = (d) adalah probabilitas dari dokumen d.

Untuk menghitung probabilitas P(d|c), digunakan rumus sebagai berikut :

$$P\big((d|c)\big) = \prod (P(t|c)\text{^}nt$$

Keterangan :

P = $\left(\frac{t}{c}\right)$ adalah probabilitas kemunculan t (kata atau istilah) dalam kategori c.

Nt adalah jumlah kemunculan fitur t dalam dokumen d.

Untuk menghitung probabilitas P(c) dan P(d), digunakan rumus sebagai berikut :

P(c) = $\left(\frac{Nc}{N}\right)$

Keterangan :

Nc adalah jumlah dokumen dalam kategori c.

N adalah jumlah total dokumen.

P (d) dapat dihitung dengan cara yang sama seperti menghitung P (d|c), dengan asumsi bahwa dokumen tersebut termasuk dalam semua kategori yang ada. Setelah semua probabilitas diperoleh, dokumen dapat diklasifikasikan ke dalam kategori dengan probabilitas tertinggi. Kategori dengan probabilitas tertinggi dianggap sebagai kategori dokumen yang paling mungkin.

2.7. Evaluasi Model

Langkah terakhir dalam penelitian ini, yaitu evaluasi, apakah algoritma Naive Bayes merupakan pengklasifikasi teks yang baik sehingga dapat memperoleh nilai akurasi yang tinggi pada analisis sentimen pengguna terhadap aplikasi Quora. Proses evaluasi dilakukan menggunakan confusion matrix dengan melihat nilai akurasi dan ROC dari hasil eksperimen yang dilakukan [6].

3. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan pengujian sentimen analisis terhadap opini pengguna terkait aplikasi Quora pada Google Play Store. Data yang diolah menggunakan bahasa pemrograman python dan jupyter notebook sebagai text editor. Data yang diperoleh berjumlah 1000 data. Dilakukan data *processing*, yaitu proses case *case folding,* remove punctuation, tokenizing, stopwords removal, dan *lemmatization.* Selanjutnya data diberikan label positif, netral dan negatif. Data dibagi menjadi 2, yaitu data training dan data testing dengan perbandingan data 80:20. Data training berjumlah 800 data dan data testing berjumlah 200 data. Selanjutnya dilakukan pembobotan TF-IDF.

Tahap selanjutnya, yaitu penerapan algoritma Naive Bayes terhadap data yang telah selesai melalui proses preprocessing dan pembobotan. Proses evaluasi dalam

 ISSN 2614-8277

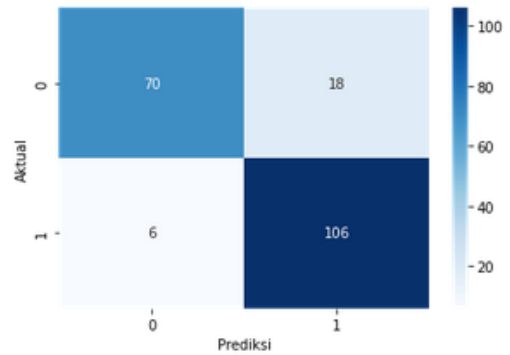
penelitian ini menggunakan Confusion Matrix yang berfungsi untuk memvisualisasikan hasil kinerja dari suatu algoritma [6].

```
# Melakukan prediksi pada data uji
prediksi = model.predict(X_test)

# Evaluasi model, misalnya dengan akurasi
from sklearn.metrics import accuracy_score

akurasi = accuracy_score(y_test, prediksi)
print("Akurasi:", akurasi)
```

```
Akurasi: 0.88
```

Gambar 1. Evaluasi model

Dari hasil uji coba diperoleh akurasi sebesar 0.88 dengan nilai True Negatif (data negatif yang diprediksi benar) sebanyak 70 data, False Positif (data negatif namun diprediksi sebagai data positif) sebanyak 18 data, False Negatif (data positif namun diprediksi sebagai data negatif) sebanyak 6 data, dan True Positif (data positif yang diprediksi benar) sebanyak 106 data. Dapat dilihat pada Gambar 2.



Gambar 2. Confusion Matrix

Berdasarkan confusion matriks yang dihasilkan maka dapat dilihat hasil dari nilai akurasi yang didapatkan. Hasil akurasi yang didapatkan, yaitu 0.88, diperoleh hasil akurasi yang tinggi karena False Positif dan False Negatif memiliki nilai yang lebih kecil dari nilai True Negatif dan True Positif. Hasil akurasi dapat di jabarkan dalam bentuk perhitungan menggunakan persamaan dibawah ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} X100\%$$

$$Accuracy = \frac{106 + 70}{106 + 70 + 18 + 6} X100\%$$

$$Accurasy = \frac{176}{200} X100\% = 88\%$$

4. KESIMPULAN

Penelitian ini menemukan bahwa algoritma Naive Bayes memiliki kinerja yang baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Quora di Play Store, dengan akurasi mencapai 88%. Temuan ini dapat membantu pengembang dalam mengidentifikasi aspek positif atau negatif yang dianggap oleh pengguna, sehingga mereka dapat membuat keputusan yang tepat untuk meningkatkan pengalaman pengguna. Penelitian selanjutnya direkomendasikan untuk meningkatkan akurasi, menggunakan seleksi fitur, dan menambah jumlah data untuk hasil yang lebih baik.

5. REFERENSI

[1] https://play.google.com/store/apps/details?id=com.quora.android&hl=id&gl=US&pli=1. (diakses pada 06 Mei 2023)

[2] I. G. C. Purnama, N. A. Sanjaya ER, and L. A. A. R.Putri,"Sentiment Analysis of Snack Review Using the Naïve Bayes Method," JELIKU (Jurnal Elektron. Ilmu Komput. Udayana), vol. 8, no. 3, p. 333, 2020, doi: 10.24843/jlk.2020.v08.i03.p16.

[3] A. V. Sudiantoro and E. Zuliarso, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier ANALISIS SENTIMEN TWITTER MENGGUNAKAN TEXT MINING DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER," vol. 10, no. 2, pp. 69–73, 2018.

[4] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," pp. 8–9, 2018

[5] L. Ardiani and H. Sujaini, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak

Implementation of Sentiment Analysis of Community Responses to Development in Pontianak City," vol. 8, no. 2, pp. 44–51, 2020.

[6] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH", JAMIKA, vol. 13, no. 1, pp. 84-93, Apr. 2023.

[7] L. M. Chaira, N. Hidayat, I. L. Wibowo, and I. Mukhlash, "Pemilihan Jenis Asuransi Berdasarkan Demografi Calon Pemegang Polis Dengan Metode Naive Bayes Classifier," J. Math. Its Appl., Vol. 13, pp. 11–20, 2016, doi: https://doi.org/10.12962/j1829605X.v13i2.1931.

[8] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," SMATIKA, vol. 10, no. 2, pp. 71–76, 2020

[9] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1, 2021.

[10] "1.9. Naive Bayes." https://scikit-learn.org/stable/modules/naive_bayes.html (accessed Jun. 07, 2023).