Jusikom: Jurnal sistem informasi Ilmu Komputer
Vol. 9 No.1 Agustus 2026, e-issn: 280-2879

Klasifikasi sentimen analisis aplikasi Jaket boat pada ulasan playstore menggunakan algoritma naïve bayes

Aditya Yogi Pratama[1], Baenil Huda[2], Fitria Nurapriani[3]
[1,2,3]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang
E-mail : si21.adityapratama@mhs.ubpkarawang.ac.id

Abstrak

Kemajuan teknologi informasi telah mendorong penggunaan aplikasi transportasi digital, termasuk Jaket Boat. Seiring bertambahnya user, review di Google Play Store menjadi sumber data penting bagi pengembang dalam meningkatkan layanan. Namun, analisis manual terhadap ulasan dalam jumlah besar kurang efisien. Tujuan penelitian ini adalah mengklasifikasikan review pengguna dengan algoritma Naive Bayes. Data dikumpulkan melalui web scraping dan diproses dengan Natural Language Processing (NLP). Proses pre-processing melibatkan pembersihan teks, case folding, tokenisasi, stopword removal, dan stemming. Data yang telah diproses kemudian dikonversi ke bentuk numerik dengan metode TF-IDF (Term Frequency-Inverse Document Frequency) sebelum diklasifikasikan menggunakan model Naive Bayes. Evaluasi model dilakukan dengan menggunakan metrik akurasi, recall, precision, dan F1-score. Penelitian ini menunjukan hasil bahwa algoritma Naive Bayes mampu mengklasifikasikan sentimen ulasan dan akurasi yang didapat yang cukup tinggi. Dengan demikian, penelitian ini dapat membantu pengembang aplikasi dalam memahami permasalahan yang dialami pengguna dan meningkatkan kualitas layanan berdasarkan hasil sentimen analisis.

Kata kunci: Klasifikasi, Naive bayes, Ulasan pengguna, NLP, Aplikasi Jaket Boat.

Abstract

Advances in information technology have encouraged the use of digital transportation applications, including Jaket Boat. As users grow, reviews on the Google Play Store become an important source of data for developers in improving services. However, manual analysis of a large number of reviews is less efficient. This research aims to classify the sentiment of user reviews using Naïve Bayes algorithm. Data was collected through web scraping and processed using Natural Language Processing (NLP). The preprocessing process involves text cleaning, case folding, tokenization, stopword removal, and stemming. The processed data is then converted into numerical form using the Term Frequency-Inverse Document Frequency (TF-IDF) method before being classified using the Naïve Bayes model. Model evaluation is done using accuracy, precision, recall, and F1-score metrics. The results show that the Naïve Bayes algorithm is able to classify the sentiment of reviews with a fairly high accuracy. Thus, this research can help application developers understand the problems experienced by users and improve service quality based on the results of sentiment analysis.

Keywords: Classification, Naïve Bayes, User Reviews, NLP, Boat Jacket Application.

1. PENDAHULUAN

Saat ini, penggunaan smartphone berkembang sangat pesat di negara Indonesia, dengan jumlah penduduk yaitu 250 juta jiwa, menjadikannya pasar yang signifikan bagi pengguna smartphone. Menurut Emarketer, sebuah lembaga pemasaran digital, tahun 2018 ada lebih dari 100 juta pengguna smartphone. Dengan angka tersebut, indonesia akan menduduki peringkat keempat di dunia dalam hal pengguna smartphone aktif, di belakang Amerika Serikat, Cina, dan India [1]. Teknologi yang pesat sudah membawa perubahan drastis di berbagai aspek kehidupan, termasuk dalam dunia transportasi. Aplikasi berbasis digital semakin banyak digunakan untuk memudahkan mobilitas masyarakat. Salah satu layanan transportasi yang berkembang adalah aplikasi Jaket Boat, yang menawarkan kemudahan dalam pemesanan transportasi air secara daring. Seiring dengan meningkatnya jumlah pengguna, berbagai ulasan dan masukan dari pengguna mengenai pengalaman mereka terhadap aplikasi ini dapat ditemukan di platform seperti Google Play Store. Scoring dan ulasan adalah fitur yang sangat menonjol di playstore, yang memungkinkan pengguna untuk memberikan timbal balik tentang aplikasi yang mereka gunakan [2]. Ulasan di Play Store adalah informasi yang sangat berharga untuk pengembang aplikasi dalam mengevaluasi kualitas layanan dan meningkatkan fitur yang ada. Namun, jumlah ulasan yang besar membuat sulit bagi pihak pengembang untuk menganalisis manual satu per satu. Maka dari itu, perlu metode yang bisa secara otomatis mengklasifikasikan sentimen ke dalam kategori negatif, netral atau positif.

Metode lgoritma Naïve Bayes Classifier adalah klasifikasi yang digunakan dalam studi analisis sentimen ini. Python adalah bahasa pemrograman yang digunakan dan Google Collab sebagai alat yang digunakan. Keuntungan ketika menggunakan Naive Bayes adalah bahwa ia hanya perlu sedikit data pelatihan untuk mengidentifikasi skala pengukuran yang diperlukan untuk proses klasifikasi [3]. Metode yang cukup sering dipakai dalam sentimen analisis adalah algoritma Naive Bayes. Naive Bayes merupakan metode klasifikasi berbasis probabilitas yang sering digunakan dalam pengolahan bahasa alami karena kemampuannya dalam menangani data teks dengan baik dan efisien. Dengan diterapkan nya metode ini analisis sentimen terhadap ulasan pengguna aplikasi Jaket Boat dapat dilakukan secara otomatis dan sistematis, dan dapat membantu pengembang mengambil keputusan yang lebih baik untuk meningkatkan layanan aplikasi. Selain itu, penerapan analisis sentimen juga memiliki manfaat bagi pengguna aplikasi, terutama bagi calon pengguna yang ingin mengetahui kualitas aplikasi sebelum mengunduh dan menggunakannya. Dengan adanya klasifikasi sentimen, calon pengguna dapat memperoleh gambaran umum mengenai pengalaman pengguna lain terhadap aplikasi Jaket Boat. Hal ini diharapkan dapat meningkatkan transparansi serta kepuasan pengguna terhadap layanan yang diberikan.

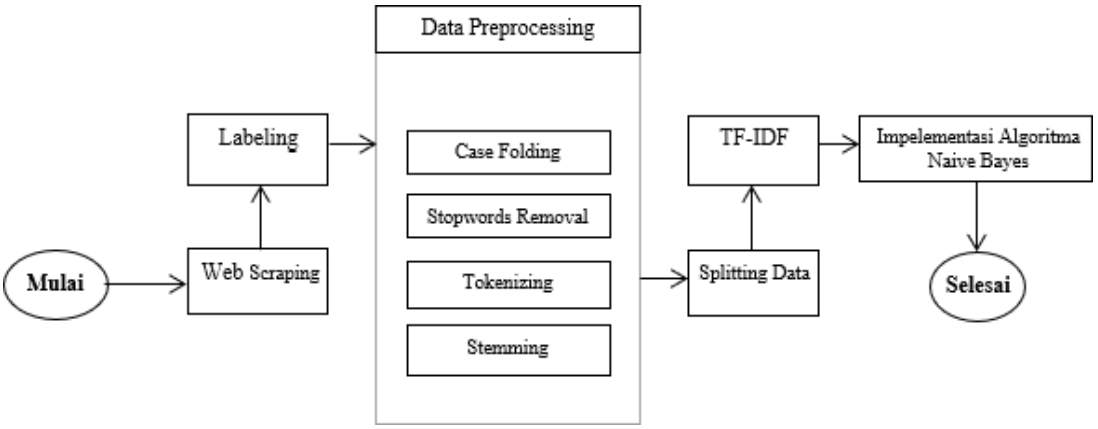
Berdasarkan penelitian yang dilakukan oleh Irfandi Analisis Sentimen Opini Masyarakat Mengenai Jasa Ekspedisi Anteraja dengan menggunakan metode Naive Bayes. Penelitian ini membahas permasalahan yang dihadapi perusahaan ekspedisi terkait pelayanan customer service serta berbagai keluhan yang disampaikan oleh para pengguna jasa. Layanan Anteraja mendapatkan beragam tanggapan dari pengguna, sehingga penting untuk mengetahui bagaimana persepsi masyarakat terhadap kualitas pelayanannya. Dalam penelitian ini digunakan algoritma Naive Bayes sebagai metode analisis, dengan data yang dikumpulkan dari platform media sosial Twitter sebanyak 1180 data. Hasil pengujian dengan algoritma tersebut menunjukkan tingkat akurasi sebesar 85,06%.[4]. Berkaitan dengan isu perpindahan orang ke perangkat digital untuk konsumsi musik dan menawarkan informasi mengenai kebutuhan dan preferensi pengguna untuk menikmati aplikasi Spotify. Sebanyak 3600 ulasan aplikasi Spotify di Google Play dikumpulkan untuk penelitian ini. Selain itu, algoritma Naive Bayes, yang memiliki hasil akurasi kinerja vektor sebesar 74,85% juga digunakan [5]

Set pelatihan dan set pengujian adalah dua set data ulasan. Set uji akan digunakan untuk berkontribusi secara independen pada penilaian akhir, sedangkan set latih akan digunakan

untuk melatih model Naive Bayes Classifier. Telah dibuktikan bahwa pendekatan ini menggunakan daya komputasi yang lebih kecil daripada pengklasifikasi teks lainnya. Salah satu manfaat dari pendekatan ini adalah, bahkan ketika digunakan pada database yang sangat besar, pendekatan ini dapat meramalkan probabilitas dengan cepat dan akurat. Penelitian ini akan membahas penerapan algoritma Naive Bayes dalam klasifikasi sentimen pengguna aplikasi Jaket Boat di Play Store. Penelitian ini diharapkan bisa memberikan dampak untuk pengembangan aplikasi berbasis digital, khususnya dalam meningkatkan kualitas layanan transportasi berbasis aplikasi daring. Selain itu, juga dapat dijadikan referensi bagi pengembang aplikasi lain dalam mengimplementasikan analisis sentimen untuk meningkatkan layanan mereka.

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah naive bayes dengan pendekatan kuantitatif. Penelitian ini akan melakukan pengumpulan, pemrosesan, dan analisis data ulasan user aplikasi Jaket Boat dengan metode Naive Bayes untuk klasifikasi sentimen.



Gambar 2.1 Tahapan Penelitian

2.1 Pengumpulan Data

Pengambilan data dilakukan dengan metode web scraping pada salah satu aplikasi booking transportasi kapal yaitu jaket boat (https://www.kapaltiket.com), khususnya review mengenai permasalahan aplikasi saat ini. Data didapatkan dari review pengguna aplikasi Jaket Boat di Play Store. Proses scraping dilakukan untuk mengambil teks ulasan secara otomatis.

2.2 Pelabelan

Pelabelan adalah proses pemberian label pada data yang sudah dikumpulkan. Dalam penelitian ini, label sangat penting untuk mengklasifikasikan data sesuai dengan sentimen yang diungkapkan di dalamnya. Data akan diberi label sesuai dengan peringkat setelah dikumpulkan. Dua kategori-positif dan negatif-kemudian diberikan pada kumpulan data yang dikumpulkan untuk setiap ulasan [6]

2.3 Pre-processing Data

Seluruh ulasan dari dataset disaring selama tahap ini guna menghilangkan kata-kata yang tidak akan mempengaruhi klasifikasi. Untuk mengurangi volume kosakata, preprocessing berupaya menghilangkan noise dan menstandarkan bentuk kata. Beberapa tugas, termasuk clenasing, eliminasi stopword, case folding, tokenizing, filtering, dan stemming, harus diselesaikan pada tahap ini.

1. Pembersihan Data : menghapus karakter khusus, tanda baca yang tidak diperlukan dan angka.
2. Tokenisasi : memecah teks menjadi kata-kata individual.
3. Stopword Removal : menghapus kata-kata umum yang tidak memiliki makna penting.
4. Stemming : mengubah kata menjadi bentuk dasar menggunakan algoritma stemming.

2.4 Transformation

Tujuan utama dari transformasi data adalah guna mengubah skala pengukuran data asli ke dalam format yang berbeda agar data tersebut sama dengan anggapan yang mendasari analisis teknik data mining yang berbeda [7]. Dengan menentukan bobot setiap kata, prosedur transformasi data disebut Term Frequency-Inverse Document Frequency.

2.5 Ekstrasi Fitur TF-IDF

Ekstraksi fitur dilakukan menggunakan metode Term Frequency-Inverse Document Frequency untuk merepresentasikan teks dalam bentuk vektor numerik. TF-IDF menghitung bobot tiap kata dalam dokumen berdasarkan frekuensi kemunculannya serta keunikan nya.

2.5.1 Prepocessing Teks

Menghapus karakter khusus, angka, dan tanda baca. Mengubah semua teks menjadi huruf kecil (lowercasing). Menghapus kata-kata umum yang tidak memiliki makna signifikan (stopwords removal). Melakukan stemming atau lemmatization untuk mengembalikan kata ke bentuk dasarnya.

2.5.2 Perhitungan TF IDF

Term Frequency (TF) dihitung sebagai jumlah kemunculan kata dalam suatu dokumen, dinormalisasi dengan panjang dokumen. Inverse Document Frequency dihitung berdasarkan logaritma terbalik dari jumlah dokumen yang mengandung kata tertentu, sehingga kata yang lebih jarang muncul memiliki bobot lebih tinggi. Bobot akhir TF-IDF diperoleh dengan mengalikan nilai TF dan IDF untuk setiap kata dalam dokumen.

2.5.3 Pembentukan Vektor Fitur

Hasil perhitungan TF-IDF digunakan untuk membentuk representasi dokumen dalam bentuk vektor. Setiap dokumen direpresentasikan sebagai vektor ber dimensi n, dimana n merupakan jumlah kata unik dalam seluruh korpus setelah preprocessing.

2.6 Algoritma Naïve Bayes

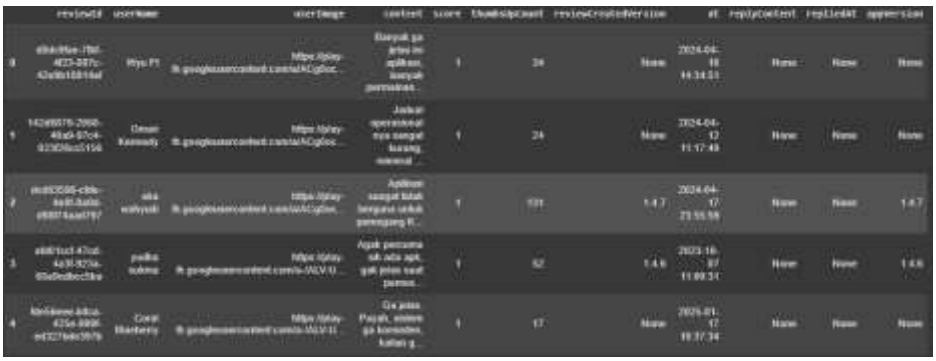
Proses menemukan pengetahuan dalam data dalam jumlah besar disebut data mining. Mengeksplorasi pengetahuan dalam data yang sangat besar dikenal sebagai data mining. Untuk menemukan informasi berharga dalam kumpulan data yang besar, teknik-teknik dari matematika digunakan data mining, lalu statistik, kecerdasan buatan, dan pembelajaran mesin. Algoritma yang paling banyak digunakan data mining salah satunya yaitu algoritma Naive Bayes, yang beroperasi di bawah premis dasar bahwa setiap peristiwa atau kondisi tidak

bergantung pada yang lain. Anggapan ini memungkinkan Naive Bayes untuk mengklasifikasikan data teks secara akurat dan cepat, terutama ketika data berisi sejumlah besar fitur atau kategori [8].

3. HASIL PEMBAHASAN

3.1 Data Selection

Metode scraping digunakan di penelitian ini dan data di ambil dari ulasan aplikasi jaket boat yang terdapat di google play store dengan jumlah 616 ulasan dan bahasa Pemrograman Python pada Google Colaboratory. Hasilnya akan diekspor dalam format .csv setelah proses penggalian data selesai. Gambar 3.1 di bawah ini memberikan contoh data ulasan yang diperoleh :



Gambar 3.1 Hasil Scrapping Data

Setelah scraping data, data awal yang tidak penting disortir untuk menghapus semua kecuali kolom nama, ulasan, dan skor (peringkat).

	userName	score	content
0	Wyu Ff	1	Banyak ga jelas ini aplikasi, banyak permainan...
1	Omari Kennedy	1	Jadual operasional nya sangat kurang, minimal ...
2	eka wahyudi	1	Aplikasi sangat tidak berguna untuk pemegang K...
3	yudha sukma	1	Agak percuma sih ada apk, gak jelas saat pemes...
4	Coral Blueberry	1	Ga jelas. Payah, sistem ga konsisten, kalian g...

Gambar 3.2 Data Berurutan

3.2 Labeling

Memberi label pada setiap sentimen sesuai dengan peringkat yang diperoleh dari temuan penggalian data adalah hal berikutnya setelah data selesai digali. yang diberikan oleh peninjau. Jika bintangnya antara empat dan lima, pelabelan akan menjadi positif; jika kurang dari tiga, pelabelan akan menjadi negatif. Prosedur pelabelan diilustrasikan pada Gambar 3.3 di bawah ini :

```
def pelabelan(score):
  if score < 3:
    return 'Negatif'
  elif score == 4 :
    return 'Positif'
  elif score == 5 :
    return 'Positif'
my_df['Label'] = my_df ['score'].apply(pelabelan)
my_df.head(50)
```

Gambar 3.3 Tampilan Proses Labeling

Setelah eksekusi kode di Gambar 3.3 berhasil, Gambar 3.4 menampilkan temuan pelabelan sentimen sebagai berikut :

	content	score	Label
kasih bintang 1 pun kebagusan, Ga Guna		1	Negatif
aplikasi pemerintahan ini?, apa ini app nya yg...		2	Negatif
tiap cari tiket gak pernah ada buat apa coba a...		1	Negatif
aplikasinya masih banyak banget kurangnya 1. p...		1	Negatif
aplikasi ga fear banyak permainan dibelakang ,...		1	Negatif
Kacaw sistem war tiketnya, banyak titipan orda...		1	Negatif

Gambar 3.4 Hasil Labeling

3.3 Data Preprocessing

Selanjutnya adalah tahapan text processing dimana data kemudian akan diproses untuk menghilangkan kata-kata yang tidak diperlukan, adapun tahapan dari preprocessing tersebut sebagai berikut:

3.3.1 Cleansing

Cleansing merupakan tahapan penting dalam pengolahan data yang bertujuan untuk menghapus informasi yang tidak berguna, salah, tidak sesuai konteks, duplikat, atau tidak konsisten, karena elemen-elemen tersebut dapat menurunkan kualitas dan keakuratan hasil analisis.

```
content    0
score      0
Label     16
dtype: int64
```

Gambar 3.5 data kosong

```
content    0
score      0
Label      0
dtype: int64
```

Gambar 3.6 Data bersih

3.3.2 Case Folding

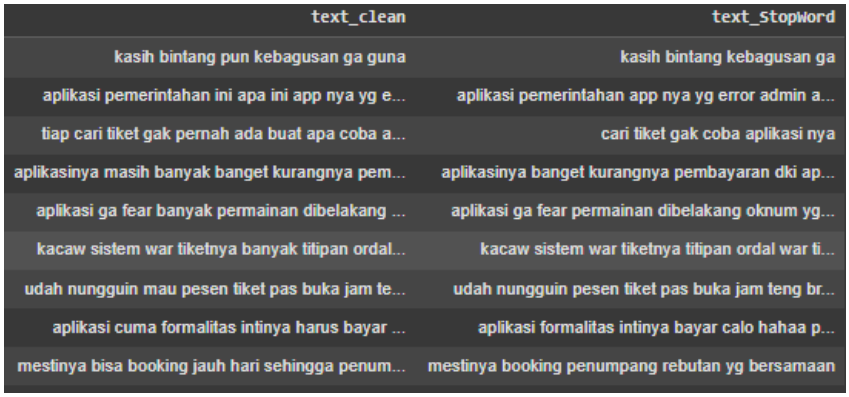
Case folding adalah mengubah bentuk keseluruhan huruf untuk membuatnya lebih kecil. Gambar berikut ini memberikan hasil penerapannya :

	content	score	Label	text_clean
0	kasih bintang 1 pun kebagusan, Ga Guna	1	Negatif	kasih bintang pun kebagusan ga guna
1	aplikasi pemerintahan ini?, apa ini app nya yg...	2	Negatif	aplikasi pemerintahan ini apa ini app nya yg e...
2	tiap cari tiket gak pernah ada buat apa coba a...	1	Negatif	tiap cari tiket gak pernah ada buat apa coba a...
3	aplikasinya masih banyak banget kurangnya 1. p...	1	Negatif	aplikasinya masih banyak banget kurangnya pem...
4	aplikasi ga fear banyak permainan dibelakang ...	1	Negatif	aplikasi ga fear banyak permainan dibelakang ...
5	Kacaw sistem war tiketnya, banyak titipan ordal...	1	Negatif	kacaw sistem war tiketnya banyak titipan ordal...
6	udah nungguin mau pesen tiket pas buka jam 5 t...	1	Negatif	udah nungguin mau pesen tiket pas buka jam te...
7	aplikasi cuma formalitas , intinya harus bayar ...	1	Negatif	aplikasi cuma formalitas intinya harus bayar ...
8	mestinya bisa booking jauh hari, sehingga penu...	1	Negatif	mestinya bisa booking jauh hari sehingga penum...
9	sangat tidak membantu	1	Negatif	sangat tidak membantu

Gambar 3.7 Hasil case folding

3.3.3 Stopword Remooval

Metode filtering atau stop removal merupakan analisis data untuk tahap setelah case folding untuk menjamin kualitas data yang digunakan. Tujuan menghilangkan kata sambung adalah untuk membuat kalimat ulasan jadi lebih pendek dan mudah dikelola.

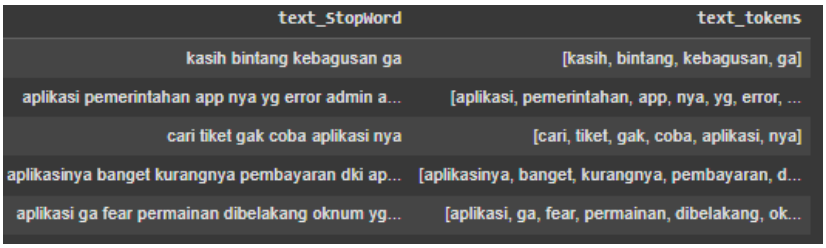


text_clean	text_StopWord
kasih bintang pun kebagusan ga guna	kasih bintang kebagusan ga
aplikasi pemerintahan ini apa ini app nya yg e...	aplikasi pemerintahan app nya yg error admin a...
tiap cari tiket gak pernah ada buat apa coba a...	cari tiket gak coba aplikasi nya
aplikasinya masih banyak banget kurangnya pem...	aplikasinya banget kurangnya pembayaran dki ap...
aplikasi ga fear banyak permainan dibelakang ...	aplikasi ga fear permainan dibelakang oknum yg...
kacaw sistem war tiketnya banyak titipan ordal...	kacaw sistem war tiketnya titipan ordal war ti...
udah nungguin mau pesen tiket pas buka jam te...	udah nungguin pesen tiket pas buka jam teng br...
aplikasi cuma formalitas intinya harus bayar ...	aplikasi formalitas intinya bayar calo hahaa p...
mestinya bisa booking jauh hari sehingga penum...	mestinya booking penumpang rebutan yg bersamaan

Gambar 3.8 Hasil stopword removal

3.3.4 Tokenisasi

Proses tokenisasi melibatkan pembagian frasa atau teks data menjadi kata-kata atau token.



text_StopWord	text_tokens
kasih bintang kebagusan ga	[kasih, bintang, kebagusan, ga]
aplikasi pemerintahan app nya yg error admin a...	[aplikasi, pemerintahan, app, nya, yg, error, ...
cari tiket gak coba aplikasi nya	[cari, tiket, gak, coba, aplikasi, nya]
aplikasinya banget kurangnya pembayaran dki ap...	[aplikasinya, banget, kurangnya, pembayaran, d...
aplikasi ga fear permainan dibelakang oknum yg...	[aplikasi, ga, fear, permainan, dibelakang, ok...

Gambar 3.9 Hasil Tokenisasi

3.3.5 Stemming

Proses stemming melibatkan pengurangan kata menjadi bentuk yang paling dasar yaitu menghilangkan seperti imbuhan prefiks, sufiks, dan infiks, agar menghasilkan kata dasar dengan arti yang sama.

```
1 : kasih : kasih
2 : bintang : bintang
3 : kebagusan : bagus
4 : ga : ga
5 : aplikasi : aplikasi
6 : pemerintahan : perintah
```

Gambar 3.10 Hasil Stemming

3.4 TF IDF

Teknik pembobotan Term Frequency-Inverse Document Frequency digunakan dalam proses transformasi. Setiap kata dalam ulasan berdasarkan perhitungan yang dilakukan dengan menggunakan metode TF-IDF. Ada dua tahap yang digunakan dalam penerapannya, yaitu count vectorizer dan TF-IDF vectorizer

```
from sklearn.feature_extraction.text import CountVectorizer

vectorizer = CountVectorizer()
vectorizer.fit(X_train)

CountVectorizer
CountVectorizer()

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

Gambar 3.11 Tahapan TF-IDF

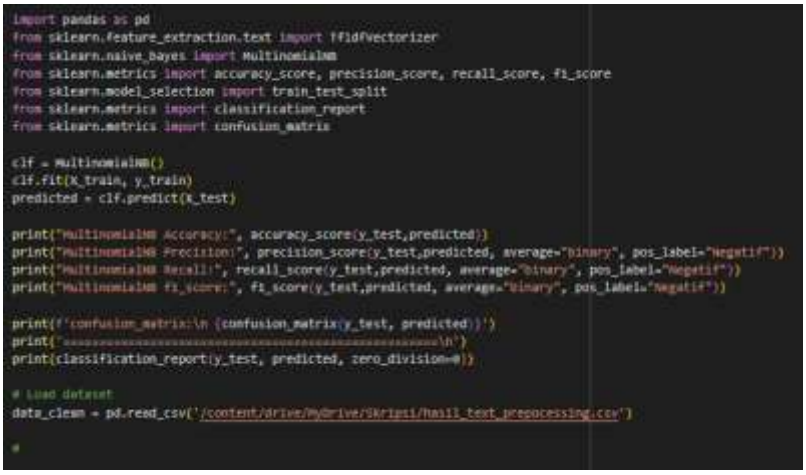
```
[ ] from sklearn.feature_extraction.text import TfidfVectorizer

    tfidf_vectorizer = TfidfVectorizer()
    tfidf_train = tfidf_vectorizer.fit_transform(X_train)
    tfidf_test = tfidf_vectorizer.transform(X_test)

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

    (480,)
    (480,)
    (120,)
    (120,)
```

Gambar 3.12 Tahapan Count vectorizer

Berdasarkan jumlah kata, potongan kode pada Gambar 3.12 akan menyiapkan data teks dengan mengubahnya menjadi representasi numerik.

3.5 Algoritma Naive Bayes

Algoritma ini digunakan untuk klasifikasi, yang dilakukan setelah tahap transformasi TF-IDF. Dengan menggunakan sumber daya Google Collaboratory, pemrograman Python digunakan untuk melakukan perhitungan presisi, recall, dan presisi berdasarkan hasil klasifikasi Naive Bayes. Sebelum klasifikasi, data yang telah diproses sebelumnya diberi label berdasarkan masukan dari user, lalu dibagi dengan 1.000 data valid menjadi 80% pelatihan dan 20% pengujian.



```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix

clf = MultinomialNB()
clf.fit(X_train, y_train)
predicted = clf.predict(X_test)

print("MultinomialNB Accuracy:", accuracy_score(y_test,predicted))
print("MultinomialNB Precision:", precision_score(y_test,predicted, average="binary", pos_label="Negatif"))
print("MultinomialNB Recall:", recall_score(y_test,predicted, average="binary", pos_label="Negatif"))
print("MultinomialNB f1_score:", f1_score(y_test,predicted, average="binary", pos_label="Negatif"))

print(f'confusion_matrix:\n {confusion_matrix(y_test, predicted)}')
print('==================================================\n')
print(classification_report(y_test, predicted, zero_division=0))

# Load dataset
data_clean = pd.read_csv('/content/drive/MyDrive/Skripsi/Hasil_text_preprocessing.csv')

#
```

Gambar 3.12 Penerapan Algoritma Naïve Bayes

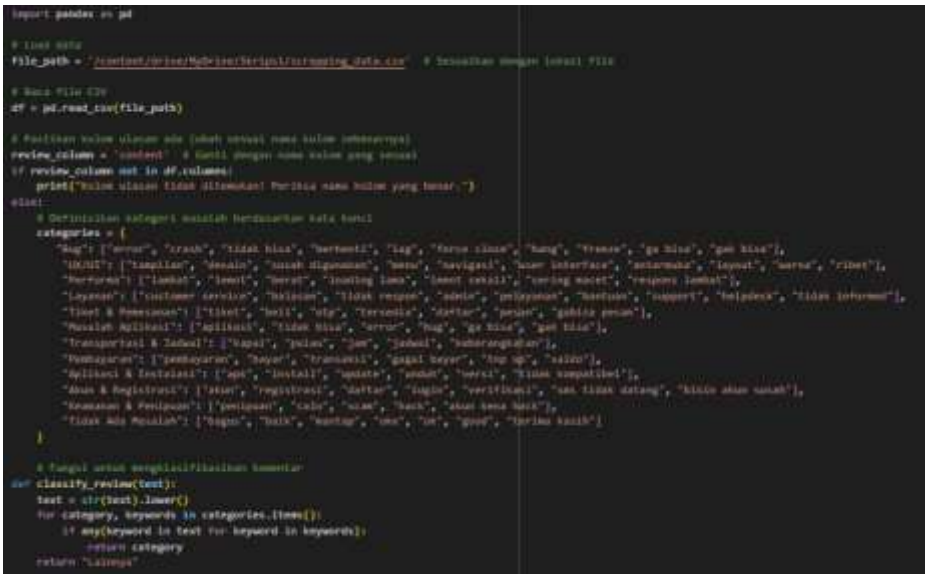
```
MultinomialNB Accuracy: 0.9083333333333333
MultinomialNB Precision: 0.907563025210084
MultinomialNB Recall: 1.0
MultinomialNB f1_score: 0.9515418502202643
confusion_matrix:
 [[108   0]
 [ 11   1]]
==================================================

              precision    recall  f1-score   support

     Negatif       0.91      1.00      0.95       108
     Positif       1.00      0.08      0.15        12

    accuracy                           0.91       120
   macro avg       0.95      0.54      0.55       120
weighted avg       0.92      0.91      0.87       120
```

Gambar 3.13 Hasil Klasifikasi (akurasi, presisi, recall dan f1-score)

Mengacu pada Gambar 3.13, proses klasifikasi dengan algoritma Naive Bayes dilakukan dengan pembagian data sebesar 80% sebagai data latih dan 20% sebagai data uji. Hasil dari proses ini menunjukkan tingkat akurasi sebesar 91%. Nilai presisi yang diperoleh mencapai 91% pada kelas negatif dan 100% pada kelas positif. Untuk nilai recall, kelas negatif memperoleh 100%, sementara kelas positif hanya sebesar 8%. Selanjutnya, nilai f1-score tercatat sebesar 95% untuk kelas negatif dan 15% untuk kelas positif.

3.6 Kategori Masalah Ulasan

Setelah mengetahui klasifikasi ulasan, selanjutnya kategorikan setiap ulasan kedalam kategori masalah yang ada seperti Bug, ui/ux, Performa, Layanan, Tiket & Pemesanan, Masalah Aplikasi, Pembayaran, Aun registrasi, Keamanan, dan Tidak ada masalah.



Gambar 3.14 Tahapan Pengkategorian Masalah Aplikasi

```
Distribusi Kategori Masalah:
Kategori Masalah
Tiket & Pemesanan        38.636364
Bug                      21.103896
Masalah Aplikasi         15.584416
Lainnya                   9.415584
Tidak Ada Masalah         3.733766
UX/UI                     2.597403
Layanan                   2.110390
Transportasi & Jadwal     1.948052
Pembayaran                1.785714
Aplikasi & Instalasi      1.785714
Akun & Registrasi         0.811688
Keamanan & Penipuan       0.324675
Performa                  0.162338
Distribusi kategori disimpan di: /content/drive/MyDrive/Skripsi/hasilpersentase.csv
```

3.15 Hasil Pengkategorian masalah aplikasi

Dari hasil pengkategorian masalah di temukan bahwa permasalahan aplikasi yang sangat di keluhkan pengguna adalah tentang pemesanan tiket untuk kapal yang masih sering terkendala sebesar 38%.

4. KESIMPULAN

Penelitian ini adalah guna membangun suatu sistem yang dapat melakukan scraping ulasan data pengguna dari Play Store dan mengklasifikasikan sentimen ulasan tersebut menggunakan algoritma Naive Bayes. Berdasarkan hasil yang diperoleh, berikut adalah beberapa kesimpulan yang dapat ditarik:

1. Pengambilan data ulasan pengguna aplikasi Jaket Boat berhasil dilakukan menggunakan teknik *web scraping* berbasis Python. Proses ini menghasilkan total 616 data ulasan yang terdiri atas berbagai ekspresi sentimen pengguna terhadap fitur dan kinerja aplikasi Jaket Boat.
2. Tahapan pra-pemprosesan teks seperti case folding, tokenisasi, filtering, stemming, dan stopword removal mampu membersihkan dari karakter yang tidak relevan, sehingga data lebih siap untuk diklasifikasikan. Setelah pra-pemrosesan, data dibagi menjadi dua label sentimen: positif dan negatif.
3. Algoritma Naive Bayes Multinomial digunakan untuk melakukan klasifikasi sentimen. Berdasarkan pengujian terhadap data uji, model ini mampu mencapai akurasi sebesar 91%, precision sebesar 91%, recall 100%, dan F1-score 95%. Nilai-nilai ini menandakan bahwa algoritma bekerja dengan cukup baik dalam mengenali pola sentimen dalam ulasan teks berbahasa Indonesia.
4. Dari hasil klasifikasi, diketahui bahwa sebagian besar ulasan pengguna cenderung negatif terhadap aplikasi Jaket Boat, dengan fokus terbesar permasalahan pada tiket & pemesanan.
5. Secara keseluruhan, sistem klasifikasi sentimen yang dibangun dalam penelitian ini dapat menjadi alat bantu analisis opini pelanggan yang berguna bagi pengembang aplikasi dalam mengevaluasi dan meningkatkan kualitas layanan berdasarkan masukan pengguna secara otomatis.

5. SARAN

Berdasarkan penelitian dan kesimpulan yang sudah di dapat, berikut adalah saran yang dapat dipertimbangkan untuk pengembangan dan penelitian selanjutnya:

1. Penambahan jumlah data ulasan: Penelitian ini hanya menggunakan 616 data ulasan. Guna meningkatkan akurasi dan generalisasi model klasifikasi, disarankan agar memperbanyak jumlah data ulasan yang digunakan, baik dari waktu yang lebih panjang maupun dari berbagai versi aplikasi.
2. Penggunaan metode klasifikasi lain sebagai perbandingan: Untuk mengetahui performa yang lebih optimal, disarankan melakukan perbandingan dengan metode lain seperti Support Vector Machine, Random Forest, atau model deep learning seperti LSTM agar diperoleh pendekatan terbaik untuk klasifikasi sentimen ulasan berbahasa Indonesia.
3. Peningkatan pada tahap pra-pemrosesan: Beberapa ekspresi atau kata-kata dalam bahasa Indonesia yang bersifat ambigu, sarkastik, atau informal masih menjadi tantangan. Oleh karena itu, disarankan untuk menggunakan kamus sinonim, kamus bahasa gaul, atau model pemrosesan bahasa alami (NLP) berbasis embedding seperti Word2Vec atau BERT agar hasil klasifikasi lebih akurat.
4. Pengembangan sistem berbasis antarmuka (interface): Untuk penerapan nyata, sistem klasifikasi dapat dikembangkan dalam bentuk aplikasi berbasis web atau desktop, sehingga pengembang aplikasi Jaket Boat dapat secara langsung mengunggah data ulasan dan memperoleh analisis sentimen secara otomatis dan real-time.
5. Analisis sentimen multi-kelas: Penelitian ini hanya mengelompokkan sentimen ke dua kelas, yaitu positif dan negatif. Untuk penelitian selanjutnya, disarankan mengembangkan klasifikasi sentimen menjadi tiga kelas (positif, negatif, dan netral) agar dapat menangkap opini yang tidak ekstrem secara lebih akurat.

DAFTAR PUSTAKA

[1] Castleman, Kenneth R., 2004, *Digital Image Processing*, Vol. 1, Ed.2, Prentice Hall, New Jersey.

[2] Gonzales, R., P. 2004, *Digital Image Processing* (*Pemrosesan Citra Digital*), Vol. 1, Ed.2, diterjemahkan oleh Handayani, S., Andri Offset, Yogyakarta.

[3] Wyatt, J. C, dan Spiegelhalter, D., 1991, *Field Trials of Medical Decision-Aids: Potential Problems and Solutions*, Clayton, P. (ed.): *Proc. 15th Symposium on Computer Applications in Medical Care*, Vol 1, Ed. 2, McGraw Hill Inc, New York.

[4] Wyatt, J. C, Spiegelhalter, D, 2008, Field Trials of Medical Decision-Aids: Potential Problems and Solutions, *Proceeding of 15th Symposium on Computer Applications in Medical Care*, Washington, May 3.

[5] Prasetya, E., 2006, Case Based Reasoning untuk mengidentifikasi kerusakan bangunan, *Tesis,* Program Pasca Sarjana Ilmu Komputer, Univ. Gadjah Mada, Yogyakarta.

[6] Irvandi, B. Irawan, and O. Nurdiawan, "Naive Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok," INFOTECH J., vol. 9, no. 1, pp. 236–242, 2023, doi: 10.31949/infotech.v9i1.5322.

[7] O. Nurdiawan, A. Irma Purnamasari, and I. Ali, "Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana," J. Data Sci. dan Inform., vol. 1, no. 2, pp. 32–35, 2021.

[8] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," J. Inf. Syst. Res., vol. 4, no. 4, pp. 1136–1144, 2023.