

Algoritma Naïve Bayes untuk Analisis Sentiment Review Blibli.com di Google Play Store

Naïve Bayes Algorithm for Sentiment Analysis of Blibli.com Review on Google Play Store

¹Siti Nur Fadhilah*, ²Fandy Setyo Utomo

^{1,2}Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto
Jalan Letjend Pol. Soemarto No.126, Banyumas 53127, Jawa Tengah, Indonesia

*e-mail: 20sa1164@mhs.amikompurwokerto.ac.id

(*received*: 18 January 2024 *revised*: 18 March 2024, *accepted*: 20 March 2024)

Abstrak

Saat ini, terdapat sekitar 354 juta ponsel aktif di Indonesia. Dengan jumlah perangkat komunikasi yang sangat besar ini, Indonesia menduduki peringkat keempat sebagai negara dengan pengguna ponsel terbanyak di dunia. *E-Commerce*, sebagai bentuk transaksi *online*, memungkinkan pertukaran barang dan jasa secara digital untuk memenuhi kebutuhan sehari-hari. Penelitian ini memiliki tujuan untuk menerapkan analisis sentimen dengan menggunakan algoritma klasifikasi *Naive Bayes* sebagai metode untuk mengumpulkan pendapat pengguna. Dengan demikian, penelitian ini tidak hanya memberikan wawasan tentang kepuasan pelanggan terhadap *Blibli.com*, tetapi juga menjadi dasar untuk perbaikan potensial dalam layanan atau pengembangan fitur guna meningkatkan pengalaman berbelanja online. Secara total, algoritma *Naive Bayes* berhasil mencapai akurasi sekitar 84%, menunjukkan kemampuannya yang baik dalam mengategorikan sentimen ulasan. Ketika difokuskan pada data yang bersifat negatif, algoritma *Naive Bayes* menunjukkan presisi sekitar 79%, recall sekitar 95%, dan f1-score sekitar 86%, mengindikasikan keberhasilannya dalam mengidentifikasi serta mengklasifikasikan ulasan yang bersifat negatif dengan tingkat ketepatan dan sensitivitas yang tinggi. Di sisi data yang bersifat positif, algoritma *Naive Bayes* mencapai presisi sekitar 91%, recall sekitar 83%, dan f1-score sekitar 87%.

Kata kunci: *E-Commerce, Naïve Bayes, Blibli.com*

Abstract

Currently, there are approximately 354 million active mobile phones in Indonesia, placing the country fourth globally in terms of the highest number of mobile phone users. *E-Commerce*, as a form of online transaction, enables the digital exchange of goods and services to meet daily needs. This research aims to implement sentiment analysis using the *Naive Bayes* classification algorithm as a method to gather user opinions. Thus, the study not only provides insights into customer satisfaction with *Blibli.com* but also serves as a basis for potential improvements in services or feature development to enhance the online shopping experience. Overall, the *Naive Bayes* algorithm successfully achieved an accuracy of around 84%, demonstrating its proficiency in categorizing sentiment in reviews. When focusing on negative data, the *Naive Bayes* algorithm exhibited a precision of approximately 79%, recall of around 95%, and an f1-score of about 86%, indicating its success in identifying and classifying negative reviews with high precision and sensitivity. On the positive side, the *Naive Bayes* algorithm achieved a precision of about 91%, recall of around 83%, and an f1-score of about 87%.

Keywords: *E-Commerce, Naïve Bayes, Blibli.com*

1 Pendahuluan

Jumlah ponsel aktif di Indonesia saat ini mencapai 354 juta perangkat. Hal tersebut diungkap *Google* dalam survei terbarunya, *Think Tech, Rise of Foldables: The Next Big Thing in Smartphone*. Dengan jumlah penggunaan alat komunikasi tersebut, Indonesia menempati posisi keempat negara dengan pengguna ponsel terbanyak di dunia[1]. Hal itu pula yang membuat peluang yang sangat besar untuk memasarkan produk – produk mereka dalam aplikasi *mobile* yang sangat berguna dan bermanfaat bagi kehidupan masyarakat, salah satunya menggunakan *e-commerce*. Persaingan ketat di *e-commerce* Indonesia membuat para pengusaha harus terus melakukan inovasi dan langkah strategis untuk memenangkan pasar[2]. *E-Commerce* adalah suatu bentuk transaksi online di mana barang atau jasa dapat saling dipertukarkan untuk memenuhi kebutuhan sehari-hari secara digital. *E-commerce* dapat memfasilitasi transaksi jual-beli antara penjual dan pembeli melalui *platform* online. *E-commerce* tersedia dalam bentuk website dan aplikasi *mobile*, sehingga memungkinkan masyarakat atau pengguna dengan mudah mengaksesnya dimana saja dan kapan saja[3].

Seiring kemajuan teknologi, orang mulai berkomunikasi di jejaring sosial. Sekarang kita tinggal mengulas sesuatu yang mudah ditemukan orang lain. Tujuan audit adalah untuk memberikan pengalaman yang dimiliki orang terkait. Begitu pula pada layanan *e-commerce*[4]. Banyak orang yang melakukan review ini untuk mendapatkan opini tentang layanan belanja online ini. Jika layanan tersebut mendapat banyak ulasan positif, pembaca merasa lebih percaya diri saat melakukan transaksi pembelian dan penjualan di belanja elektronik tersebut. Begitu pula dengan, sebaliknya jika penilaian terhadap pelayanannya negatif maka pembaca akan berpikir ulang untuk berbisnis di bidang belanja elektronik. Salah satu *e-commerce* yang sedang ramai digunakan adalah *Blibli.com*. *Blibli.com* merupakan salah satu perusahaan *e-commerce* Indonesia dengan konsep belanja online seperti *mall*. *Blibli.com* yang bermoto *Big Choices Big Deals* ini menawarkan lebih dari 1,5 juta produk dari brand dan merchant terpercaya sehingga kualitasnya terjamin[5]. Melihat fenomenasat ini, *Blibli.com* memiliki lima juta pengunduh dengan rating 4,1, namun penggunanya hanya 127.199 orang[6]. Meskipun *Blibli.com* telah mencapai popularitas yang signifikan dengan lima juta pengunduh dan rating 4.1, perlu dilakukan sentimen analisis lebih lanjut untuk memahami secara mendalam bagaimana pengguna sebenarnya merespons pengalaman berbelanja online di platform ini.

Analisis sentimen adalah metode untuk mengekstrak data opini, secara otomatis memahami dan mengolah data teks untuk melihat emosi yang terkandung dalam opini tersebut[7]. Dalam konteks *Blibli.com*, analisis sentimen dapat membantu mengidentifikasi apakah umpan balik pengguna cenderung positif, negatif, atau netral. Metode ini tidak hanya memberikan wawasan tentang kepuasan pelanggan, tetapi juga dapat menjadi landasan untuk perbaikan layanan atau pengembangan fitur. Untuk mengimplementasikan analisis sentimen, salah satu pendekatan yang umum digunakan adalah algoritma klasifikasi Naive Bayes, yang dapat memprediksi sentimen berdasarkan kemungkinan probabilitas dari kata-kata kunci yang muncul dalam ulasan pengguna. Naive Bayes dapat menjadi alat yang efektif untuk *Blibli.com* dalam memahami dinamika opini pelanggan dan meningkatkan pengalaman belanja online mereka. Naive Bayes adalah metode klasifikasi probabilistik sederhana berdasarkan teorema Bayes yang mengasumsikan independensi setiap kondisi atau peristiwa[7]. Metode Naive Bayes sering digunakan dalam pengenalan pola dan klasifikasi teks, terutama dalam konteks analisis sentimen. Dengan memanfaatkan probabilitas dan asumsi independensi, Naive Bayes dapat dengan efisien mengklasifikasikan data ke dalam kategori yang tepat. Meskipun sederhana, metode ini sering memberikan hasil yang baik, terutama ketika data yang digunakan memenuhi asumsi dasar independensi. Namun, penting untuk diingat bahwa dalam kasus data yang sangat kompleks atau berkorelasi tinggi, performa Naive Bayes dapat terpengaruh.

Penelitian yang serupa juga pernah dilakukan menggunakan Algoritma Naive Bayes, yaitu yang dilakukan oleh Irma Putri Rahayu[8]. Penelitian yang berjudul “Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine” memiliki beberapa hasil yang dibahas. Hasil penelitian ini mencakup penggunaan algoritma Naive Bayes dan SVM dalam melakukan analisis sentimen terhadap program Kampus Merdeka. Ditemukan bahwa penggunaan algoritma Naive Bayes memberikan akurasi sebesar 86%, presisi 87%, dan recall 80%, sedangkan algoritma SVM kernel linear menghasilkan akurasi 93%, presisi 100%, dan recall 84%. Penelitian ini mengungkapkan adanya kebingungan dan respons positif maupun negatif dari masyarakat terkait pelaksanaan kebijakan MBKM, mendorong perlunya analisis sentimen untuk memberikan kejelasan

terkait program tersebut. Sebagai tambahan, penelitian sebelumnya juga dilakukan terkait analisis sentimen terhadap program MBKM dengan akurasi model SVM sebesar 80,75%. Hasil penelitian menunjukkan bahwa metode Naïve Bayes memberikan nilai akurasi terbaik, meskipun penelitian dilakukan dengan menggunakan dua algoritma, yaitu Naïve Bayes dan SVM. Selain itu, penelitian ini menggarisbawahi pentingnya langkah data preprocessing dalam proses klasifikasi teks, sambil menyarankan bahwa hasil penelitian dapat menjadi bahan evaluasi untuk pengembangan program MBKM ke depan. Secara keseluruhan, hasil penelitian ini memberikan pemahaman yang mendalam tentang analisis sentimen terhadap program Kampus Merdeka, menggunakan metode-metode yang efektif dan akurat, serta menyoroti pentingnya evaluasi dan pengembangan program MBKM berdasarkan umpan balik dari masyarakat.

Penelitian yang lain juga pernah dilakukan oleh Rita Apriani, dkk[2], dengan jurnal nya yang berjudul “Analisis Sentimen Dengan *Naïve Bayes* Terhadap Komentar Aplikasi Tokopedia”. Berdasarkan hasil pembahasan jurnal di atas, dapat disimpulkan bahwa penelitian ini menggunakan metode Naïve Bayes untuk mengklasifikasikan komentar pengguna aplikasi Tokopedia di Playstore ke dalam kategori positif dan negatif. Hasil analisis sentimen menunjukkan bahwa sebagian besar komentar adalah negatif, dengan banyak keluhan terkait dengan brand/nama Tokopedia dan kondisi barang yang tidak sesuai saat barang diterima. Penelitian ini memberikan manfaat bagi manajemen Tokopedia untuk menjaga tingkat kepercayaan pengguna/konsumen dan meningkatkan keuntungan sesuai target yang dibuat. Selain itu, penelitian ini juga menyarankan untuk menambahkan metode lain, seperti seleksi fitur information gain, untuk meningkatkan akurasi algoritma Naïve Bayes secara signifikan pada penelitian selanjutnya.

Penelitian terakhir dilakukan oleh Dewi Sekar Arum, dkk[9] yang berjudul “Analisis Sentimen Masyarakat Indonesia Terhadap *Sea Games 2023* Di *Twitter* Dengan Metode *Naïve Bayes*”. Berdasarkan penelitian yang dilakukan, ditemukan bahwa sebagian besar tanggapan masyarakat terhadap *SEA Games 2023* di *Twitter* memiliki sentimen positif sebesar 33,4%, diikuti oleh sentimen netral sebesar 59,1%, dan sentimen negatif sebesar 7,5%. Hasil klasifikasi tanggapan terhadap *SEA Games 2023* menunjukkan tingkat akurasi yang tinggi, dengan rasio perbandingan data training dan data testing berpengaruh terhadap kinerja algoritma. Dengan perbandingan 40:60, tingkat akurasi mencapai 92,70%. Metode *Naïve Bayes* digunakan dalam penelitian ini untuk menganalisis sentimen masyarakat Indonesia terhadap *SEA Games 2023* di *Twitter*. Metode ini memanfaatkan fitur-fitur dari data, seperti kata kunci, untuk mengklasifikasikan sentimen [object Object]. Dari visualisasi word cloud, dapat dilihat bahwa beberapa topik yang sering dibahas oleh pengguna *Twitter* tentang *SEA Games 2023* termasuk kata-kata positif seperti "juara", "raih", "timnas", "emas", dan "prestasi". Di sisi lain, kata-kata negatif seperti "balik", "insiden", "kalah", "minta", dan "gagal" juga muncul dalam tanggapan masyarakat. Dengan demikian, penelitian ini memberikan gambaran yang jelas tentang bagaimana masyarakat Indonesia merespons *SEA Games 2023* di *Twitter*, serta menunjukkan bahwa analisis sentimen dapat memberikan informasi berharga dalam memahami pandangan masyarakat terhadap suatu acara atau peristiwa.

Pada jurnal ini, dengan berdasarkan referensi di atas, penulis melakukan penelitian untuk memahami secara mendalam respons pengguna terhadap pengalaman berbelanja online di platform e-commerce *Blibli.com*. Pertumbuhan signifikan pengguna ponsel aktif di Indonesia, yang mencapai 354 juta perangkat, dan persaingan ketat di industri *e-commerce*, menandai urgensi untuk mengidentifikasi dan menganalisis sentimen pelanggan dalam konteks *Blibli.com*. Tujuan penelitian ini adalah mengimplementasikan analisis sentimen dengan menggunakan algoritma klasifikasi *Naive Bayes* sebagai metode untuk mengekstrak data opini pengguna. Dengan demikian, penelitian ini tidak hanya bertujuan memberikan wawasan tentang kepuasan pelanggan terhadap *Blibli.com*, tetapi juga menjadi dasar untuk potensial perbaikan layanan atau pengembangan fitur yang dapat meningkatkan pengalaman berbelanja online. Dengan merujuk pada penelitian terdahulu yang menggunakan metode serupa dalam konteks analisis sentimen, penelitian ini berupaya memberikan kontribusi pada pemahaman mendalam tentang dinamika opini pengguna terhadap layanan e-commerce di Indonesia, khususnya *Blibli.com*.

2 Tinjauan Literatur

E-Commerce adalah suatu sistem ekonomi yang melibatkan interaksi dinamis antara perusahaan, konsumen, dan komunitas melalui transaksi elektronik. Dalam konteks ini, perdagangan barang, layanan/jasa, dan pertukaran informasi dilakukan secara elektronik, menghubungkan berbagai entitas secara efisien.[10]. *E-commerce*, atau perdagangan elektronik, merujuk pada sistem transaksi online di mana barang atau jasa dapat diperjualbelikan secara digital. Konsep ini melibatkan interaksi dinamis antara perusahaan, konsumen, dan komunitas melalui transaksi elektronik. Dalam literatur, konsep ini juga sering dikaitkan dengan kemudahan akses, kecepatan, dan efisiensi dalam melakukan transaksi.

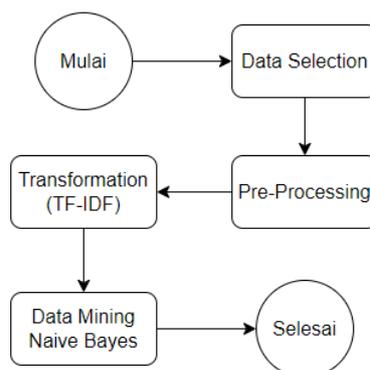
BPS telah melakukan survei e-commerce sejak tahun 2019. Survei e-commerce tahun 2023 dilakukan di 34 provinsi dan 302 provinsi/kota dengan sampel total 4.252 blok sensus dan 31.753 badan usaha[11]. Analisis sentimen adalah metode untuk mengekstrak data opini dan memahami emosi yang terkandung dalam teks. Dalam konteks *e-commerce*, analisis sentimen membantu perusahaan memahami apakah umpan balik pengguna cenderung positif, negatif, atau netral. Algoritma klasifikasi *Naive Bayes* sering digunakan untuk analisis sentimen, memprediksi sentimen berdasarkan kemungkinan probabilitas dari kata-kata kunci yang muncul dalam ulasan pengguna. Tinjauan literatur dapat mencakup kajian terdahulu yang telah mengaplikasikan metode ini dalam berbagai konteks, memberikan dasar bagi penelitian ini.

Analisis Sentimen merupakan salah satu bidang dalam penelitian text mining yang bertujuan untuk mengklasifikasikan dokumen teks. Proses ini melibatkan ekstraksi pendapat, emosi, dan evaluasi yang tertulis dari seseorang terkait suatu topik khusus. Dalam menjalankan tugasnya, Analisis Sentimen menggunakan teknik pemrosesan Bahasa alami untuk memahami dan menginterpretasikan teks dengan lebih mendalam. Secara esensial, metode ini memungkinkan identifikasi sikap positif, negatif, atau netral yang terkandung dalam suatu dokumen teks, membantu dalam memahami pandangan dan respon manusia terhadap topik tersebut. Dengan memanfaatkan kemajuan dalam pemrosesan Bahasa alami, Analisis Sentimen menjadi alat yang sangat berguna dalam mengeksplorasi opini dan perasaan yang disampaikan melalui teks, memberikan wawasan yang berharga untuk pengambilan keputusan dan pemahaman mendalam terhadap respons pengguna[12].

Algoritma klasifikasi *Naive Bayes* adalah metode sederhana namun efektif dalam mengklasifikasikan data ke dalam kategori yang tepat, terutama dalam konteks analisis sentimen. Literatur dapat membahas keunggulan dan batasan dari metode ini, serta contoh-contoh penggunaannya dalam penelitian-penelitian sebelumnya. Dengan merinci tinjauan literatur ini, penelitian ini berupaya mengintegrasikan konsep-konsep tersebut untuk memberikan pemahaman yang lebih mendalam terhadap respons pengguna terhadap layanan e-commerce, khususnya *Blibli.com*, melalui analisis sentimen dengan menggunakan algoritma klasifikasi *Naive Bayes*.

3 Metode Penelitian

Gambar 1 menggambarkan langkah-langkah penelitian untuk analisis sentimen terhadap layanan e-commerce *Blibli.com* menggunakan algoritma *Naive Bayes*.



Gambar 1 langkah – langkah penelitian

Gambar 1 merupakan Proses analisis data teks dengan menggunakan algoritma Naive Bayes dimulai dengan pemilihan data yang relevan, termasuk sumber data dan variabel yang dibutuhkan. Setelah itu, data diproses dengan tahapan preprocessing untuk membersihkan dan mempersiapkannya agar sesuai untuk analisis. Transformasi data menjadi representasi numerik menggunakan metode TF-IDF menjadi langkah kunci untuk menggambarkan bobot kata-kata dalam konteks dokumen. Tahapan ini memungkinkan penggunaan algoritma Naive Bayes, yang pada dasarnya mengklasifikasikan data berdasarkan probabilitas kata-kata kunci dalam teks. Selesai dengan tahap evaluasi hasil klasifikasi, analisis data teks ini memberikan pemahaman mendalam tentang pola dan sentimen yang terkandung dalam dataset, dan dapat memberikan dasar bagi keputusan lebih lanjut atau rekomendasi berdasarkan temuan analisis.

3.1. Data Selection

Pada tahap penyisihan ini, penulis melakukan pengumpulan informasi dan pelabelan statistik. Rekaman diambil dari utilitas *Blibli.com* di *Google Play Store* menggunakan teknik *scraping* dengan menggunakan pemrograman *Python*[13].

3.2. Pre - Processing

Data yang telah terkumpul akan melewati serangkaian tahapan preprocessing sebelum dapat dijalani analisis. Proses preprocessing ini melibatkan beberapa langkah, antara lain:

1. *Case Folding*: Mengubah seluruh kalimat menjadi huruf kecil, sehingga perbedaan kapitalisasi tidak memengaruhi analisis.
2. Penggantian Kata yang Disingkat dan Bahasa Gaul: Menyamakan kata-kata yang disingkat atau berasal dari bahasa gaul untuk memastikan konsistensi dalam data.
3. Penghapusan Huruf Berlebihan dan Tidak Bermakna: Menghilangkan huruf-huruf yang tidak memberikan makna tambahan dan bersifat berlebihan.
4. Pengubahan Bahasa yang Tidak Sesuai dengan Bahasa Indonesia: Menyesuaikan bahasa yang tidak sesuai dengan bahasa Indonesia agar data menjadi lebih homogen.
5. *Stopword Removal*: Menghapus kata-kata yang umumnya tidak mengandung makna signifikan, seperti kata hubung dan kata-kata umum lainnya.
6. Proses *Stemming*: Mengubah kata-kata menjadi bentuk dasarnya untuk mempermudah analisis data.

Langkah-langkah ini bertujuan untuk membersihkan dan mempersiapkan data sehingga dapat diolah dengan lebih efektif dan akurat selama proses analisis[14].

3.3. Transformation

Data diubah atau digabung menjadi format yang sesuai untuk diolah dalam data mining. Beberapa metode data mining memerlukan format data yang spesifik sebelum dapat diaplikasikan. Sebagai contoh, metode standar seperti analisis asosiasi dan clustering hanya dapat menerima input data kategorikal. Oleh karena itu, data dalam bentuk angka numerik yang berkelanjutan perlu diubah menjadi beberapa interval. Proses ini umumnya disebut sebagai transformasi data.

3.4. Data Mining

Data mining adalah proses analisis data yang bertujuan untuk mengidentifikasi pola atau hubungan yang signifikan dan belum diketahui sebelumnya. Dengan menggunakan metode terkini, data mining membantu mengungkap informasi yang berguna dari data, memberikan pemahaman mendalam, dan menghasilkan kesimpulan yang dapat bermanfaat bagi pemilik data[15].

4 Hasil dan Pembahasan

Penelitian ini menggunakan algoritma *Naive Bayes* yang memiliki beberapa tahapan, yaitu *data selection*, *data scrapping*, *data labeling*, *data pre-processing*, *data transformation* dan *data mining* yang kemudian akan dimasukkan algoritma terkait untuk mendapatkan hasil yang diinginkan x

4.1. Data Selection

Data selection merupakan tahapan awal dalam proses analisis data yang melibatkan pemilihan dataset atau sampel data yang akan digunakan untuk analisis lebih lanjut.

4.1.1. Data Scrapping

Data Scrapping merujuk pada proses ekstraksi informasi atau data dari suatu sumber atau situs web. Proses ini melibatkan penggunaan algoritma atau perangkat lunak khusus yang secara otomatis mengambil data dari halaman web dan menyimpannya dalam format yang dapat digunakan dan dianalisis lebih lanjut. Ada dua metode umum untuk melakukan scraping data: secara manual, di mana individu secara langsung menyalin dan menempelkan informasi dari situs web ke dalam file atau database, atau secara otomatis dengan menggunakan perangkat lunak khusus atau script yang dirancang untuk mengekstrak data dari halaman web secara otomatis[16]. Dalam penelitian ini, pengambilan data review aplikasi *Blibli.com* dilakukan secara otomatis dengan menggunakan *API* yang diambil dari *Google Play Store*. Sebanyak 1000 data hasil pengambilan tersebut kemudian disimpan dalam format *file csv*.

```
from google_play_scraper import Sort, reviews

result, continuation_token = reviews (
    'blibli.mobile.commerce',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.MOST_RELEVANT, # defaults to Sort.MOST_RELEVANT
    count=1000, # defaults to 100
    filter_score_with=None # defaults to None(means all score)
)
```

Gambar 2 proses *scrapping data*

Gambar 2, pustaka `google_play_scraper` digunakan untuk mengakses ulasan dari aplikasi *Blibli* di *Google Play Store*. Melalui fungsi `reviews`, kita menentukan aplikasi dengan mengidentifikasi ID paketnya, yaitu `'blibli.mobile.commerce'`. Selain itu, bahasa ulasan diatur ke Bahasa Indonesia (`'id'`), dan negara diatur ke Indonesia (`'id'`). Ulasan diurutkan berdasarkan relevansi dengan parameter `Sort.MOST_RELEVANT`, dan dipilih sebanyak 1000 ulasan dengan parameter `count=1000`. Tidak ada filter skor ulasan yang diterapkan (`None`), sehingga semua skor ulasan akan diambil. Hasil dari fungsi ini adalah dua nilai, yaitu `result` yang berisi daftar ulasan dan informasi terkait, serta `continuation_token` yang dapat digunakan untuk melanjutkan pengambilan ulasan jika diperlukan. Dengan konfigurasi ini, kita dapat mengumpulkan ulasan pengguna aplikasi *Blibli* dari *Google Play Store* untuk analisis lebih lanjut.

reviewId	userName	userImage	content	score	thumbUpCount	reviewCreatedVersion	at	replyContent	replyedAt	appVersion	
0	49e5e00c-7594-4824-880c-da35889d1d9c	MURID BARU	https://play.googleusercontent.com/a/ACg9oc...	Belanja nyaman, cashback mantap, gratis ongkir...	4	12	10.7.0	2024-01-12 04:04:42	Halo kak MURID BARU mantap! Makasih banyak bua...	2024-01-12 06:27:14	10.7.0
1	786152de-2f78-4cec-a52a-1b1bde622f13	Amalia Rahmawati	https://play.googleusercontent.com/a-/ALV-U...	Tampilan menu sampai pencarian hingga saat mau...	1	700	10.6.0	2023-11-08 10:13:08	Halo Kak Amalia, sangat disayangkan atas kenda...	2023-11-08 11:46:34	10.6.0
2	ad3129e7-1bd4-48f5-99de-95338e9f987	Rida Astuti	https://play.googleusercontent.com/a-/ALV-U...	Pertama kali top up tapcash di blibli. Langsung...	1	133	10.7.0	2023-12-27 13:00:33	Halo kak Rida Astuti mohon maaf terkaik... kenda...	2023-12-27 14:22:07	10.7.0
3	7c2f18ee-3e68-448a-b879-5b9a606c0789	Zy Stay	https://play.googleusercontent.com/a-/ALV-U...	Untuk cc kasih tau kepitak kalian untuk tidak...	1	294	10.7.0	2023-12-15 20:35:41	Halo kak Zy Stay mohon maaf terkaik... kenda...	2023-12-16 01:29:28	10.7.0
4	0595f5c5-3f24-47c9-baf8-5a7cbe9d384b	Jili Adisthy	https://play.googleusercontent.com/a-/ALV-U...	untuk belanja aplikasi lumayan bagus dengan pr...	2	160	10.6.0	2023-11-23 09:34:00	Halo kak Jili mohon maaf atas pengalaman belan...	2023-11-23 10:22:59	10.6.0

Gambar 3 hasil dari *scrapping data*

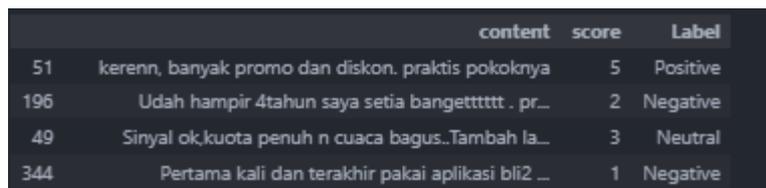
Data dari gambar 3 dapat digunakan oleh pengembang atau peneliti untuk menggali informasi lebih lanjut dari ulasan pengguna aplikasi *Blibli* di *Google Play Store*. Data tersebut dapat menjadi dasar untuk evaluasi kinerja aplikasi, pembaruan fitur, atau peningkatan layanan berdasarkan umpan balik pengguna.

4.1.2. Data Labeling

Data labeling atau Pelabelan data, juga dikenal sebagai penandaan data, adalah penetapan kategori atau label unik untuk setiap titik data atau sampel dalam kumpulan data. Tujuan dari pelabelan data adalah untuk memberikan konteks atau makna pada data sehingga dapat digunakan untuk melatih atau menguji model pembelajaran mesin[17]. Pada penelitian ini dibagi menjadi label

<http://sistemasi.ftik.unisi.ac.id>

negatif dan positif, dan proses pelabelan selanjutnya dilakukan secara otomatis menggunakan bahasa pemrograman *Python*.



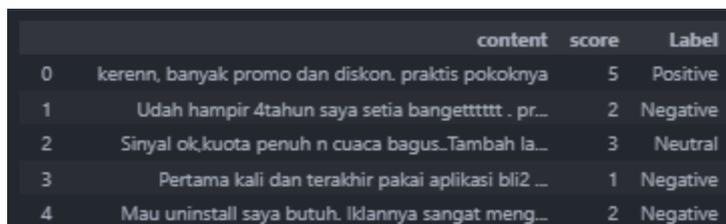
	content	score	Label
51	kerenn, banyak promo dan diskon. praktis pokoknya	5	Positive
196	Udah hampir 4tahun saya setia bangettttt . pr...	2	Negative
49	Sinyal ok,kuota penuh n cuaca bagus..Tambah la...	3	Neutral
344	Pertama kali dan terakhir pakai aplikasi bli2 ...	1	Negative

Gambar 4 hasil dari *data labeling*

Gambar 4 merupakan hasil dari *data labeling*. Hasil dari *data labeling* memainkan peran penting dalam pengembangan model analisis sentimen yang akurat dan relevan dengan konteks tertentu. Dengan menggunakan data yang telah dilabeli dengan benar, model dapat belajar dan beradaptasi untuk mengklasifikasikan sentimen dengan lebih baik.

4.2. Pre-Processing

Pre-processing data pada algoritma *Naive Bayes* merupakan serangkaian langkah yang melibatkan pembersihan data dengan penanganan nilai yang hilang, tokenisasi untuk pemisahan unit kata, pengonversian huruf menjadi huruf kecil, eliminasi kata umum, dan penerapan proses *stemming* atau *lemmatization*. Selain itu, ekstraksi fitur seperti *TF-IDF* digunakan untuk merubah data menjadi representasi vektor. Penanganan outlier dan konversi variabel kategori dengan metode *encoding* juga merupakan langkah penting dalam memastikan data siap digunakan untuk pelatihan atau prediksi menggunakan algoritma *Naive Bayes*.



	content	score	Label
0	kerenn, banyak promo dan diskon. praktis pokoknya	5	Positive
1	Udah hampir 4tahun saya setia bangettttt . pr...	2	Negative
2	Sinyal ok,kuota penuh n cuaca bagus..Tambah la...	3	Neutral
3	Pertama kali dan terakhir pakai aplikasi bli2 ...	1	Negative
4	Mau uninstall saya butuh. Iklannya sangat meng...	2	Negative

Gambar 5 hasil dari tahap *pre-processing*

Gambar 5 adalah hasil dari tahap *pre-processing*. Hasil dari langkah-langkah *pre-processing* ini adalah teks yang lebih terstruktur, terstandarisasi, dan siap untuk digunakan dalam analisis lebih lanjut. Data yang telah melalui *pre-processing* memiliki kemungkinan lebih tinggi untuk memberikan hasil analisis yang akurat dan relevan, terutama ketika digunakan dalam model pembelajaran mesin seperti algoritma klasifikasi sentimen *Naive Bayes*.

4.3. Transformation

Text transformation data pada algoritma *Naive Bayes* mengacu pada proses mengubah atau mengonversi data teks menjadi representasi yang dapat digunakan oleh algoritma tersebut untuk melakukan analisis klasifikasi. Dalam konteks ini, transformasi data melibatkan konversi teks menjadi vektor fitur yang dapat diinterpretasikan oleh *Naive Bayes*.

```
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)
```

Gambar 6 Pembobotan TF-IDF

Gambar 6 adalah pembobotan dari TF-IDF yang dilakukan. Dalam kode di atas, pustaka `scikit-learn` digunakan untuk melakukan ekstraksi fitur teks dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Objek `TfidfVectorizer` diinisialisasi untuk mengonversi teks menjadi representasi numerik. Data training (`X_train`) dan data test (`X_test`) kemudian diubah menjadi matriks TF-IDF menggunakan metode `fit_transform` dan `transform` pada objek `TfidfVectorizer`. Hasil transformasi ini menciptakan representasi numerik berdasarkan frekuensi kata-kata dalam dokumen dan invers dokumen frekuensi kata-kata tersebut. *Print statements* terakhir menampilkan dimensi dari matriks TF-IDF untuk data training dan test, serta dimensi dari label terkait (`y_train` dan `y_test`). Representasi TF-IDF yang dihasilkan dapat digunakan sebagai input untuk melatih dan menguji model pembelajaran mesin, seperti algoritma klasifikasi *Naive Bayes*, khususnya dalam tugas analisis sentimen atau pemrosesan teks lainnya.

4.4. Data Mining

Dalam proses data mining ini, hasil evaluasi menunjukkan tingkat akurasi sebesar 84%. Ketika fokus pada klasifikasi data negatif, terdapat presisi sebesar 79%, *recall* sebesar 95%, dan *f1-score* sebesar 86%. Ini mengindikasikan bahwa model mampu dengan baik mengidentifikasi data negatif, terutama dalam hal mendeteksi sebanyak mungkin data yang sebenarnya negatif. Sementara itu, pada klasifikasi data positif, presisi mencapai 91%, *recall* sebesar 83%, dan *f1-score* mencapai 87%. Hasil ini menunjukkan bahwa model efektif dalam mengenali data positif, dengan penekanan khusus pada akurasi presisi.

Perlu dicatat bahwa untuk kategori neutral, semua metrik seperti presisi, *recall*, dan *f1-score* memiliki nilai 0%. Hal ini menunjukkan bahwa model tidak memberikan hasil klasifikasi yang relevan atau bermakna untuk data kategori ini. Secara keseluruhan, meskipun terdapat keterbatasan dalam klasifikasi data neutral, model berhasil mencapai tingkat akurasi yang memuaskan, yakni sebesar 84%, dengan kinerja yang cukup baik dalam mengenali data positif dan negatif.

```
confusion_matrix:
[[91  0  5]
 [ 8  0  3]
 [16  0 77]]
=====
              precision    recall  f1-score   support

 Negative      0.79      0.95      0.86       96
  Neutral      0.00      0.00      0.00        11
  Positive      0.91      0.83      0.87       93

 accuracy              0.84       200
 macro avg           0.57      0.59      0.58       200
 weighted avg        0.80      0.84      0.82       200
```

Gambar 7 hasil data mining

Gambar 7 adalah hasil dari data mining. Hasil dari data mining pada umumnya merujuk pada temuan atau pola yang ditemukan dari data yang telah diolah menggunakan berbagai teknik analisis. Dalam konteks kode yang telah diberikan, penggunaan algoritma klasifikasi *Naive Bayes* pada matriks

TF-IDF dapat menghasilkan model yang mampu memprediksi sentimen dari ulasan pengguna. Dengan kata lain, data mining pada tahap ini dapat memberikan informasi mengenai kemungkinan sentimen positif, negatif, atau netral dari ulasan aplikasi *Blibli* di *Google Play Store*. Analisis hasil ini dapat memberikan wawasan mendalam terkait preferensi dan pengalaman pengguna, membantu pengembang atau pemilik aplikasi untuk mengidentifikasi area perbaikan atau pengembangan lebih lanjut.

5 Kesimpulan

Berdasarkan hasil evaluasi dari penelitian yang sudah dilakukan, dapat disimpulkan bahwa implementasi algoritma *Naïve Bayes* memberikan kinerja yang baik dalam menganalisis sentimen dari ulasan pengguna *Blibli.com* di *Google Play Store*. Secara keseluruhan, algoritma *Naïve Bayes* mampu mencapai tingkat akurasi sebesar 84%, menunjukkan kemampuannya dalam mengklasifikasikan sentimen ulasan dengan baik. Fokus pada data negatif, algoritma *Naïve Bayes* memiliki presisi sebesar 79%, *recall* sebesar 95%, dan *f1-score* sebesar 86%, menunjukkan keberhasilannya dalam mendeteksi dan mengklasifikasikan ulasan negatif dengan tingkat ketepatan dan sensitivitas yang tinggi. Pada sisi data positif, algoritma *Naïve Bayes* mencapai presisi sebesar 91%, *recall* sebesar 83%, dan *f1-score* sebesar 87%, menunjukkan kemampuannya dalam mengenali ulasan positif dengan akurasi dan sensitivitas yang baik.

Namun, perlu diperhatikan bahwa klasifikasi untuk ulasan netral menunjukkan keterbatasan, dengan seluruh metrik (presisi, *recall*, dan *f1-score*) mencapai nilai 0%. Hal ini menandakan bahwa model belum efektif dalam mengklasifikasikan ulasan sebagai *neutral*. Secara keseluruhan, penelitian ini menyoroti potensi algoritma *Naïve Bayes* dalam analisis sentimen ulasan *e-commerce* di platform *Google Play Store*, dengan catatan perlu pengembangan lebih lanjut, terutama dalam penanganan ulasan netral.

Referensi

- [1] S. Caroline, "Ada 354 Juta Ponsel Aktif di Indonesia, Terbanyak Nomor Empat Dunia," <https://tekno.kompas.com>. Accessed: Jan. 16, 2024. [Online]. Available: https://tekno.kompas.com/read/2023/10/19/16450037/ada-354-juta-ponsel-aktif-di-indonesia-terbanyak-nomor-empat-dunia#google_vignette
- [2] R. Apriani *et al.*, "Analisis Sentimen dengan *Naïve Bayes* Terhadap Komentar Aplikasi Tokopedia," *J. Rekayasa Teknol. Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, [Online]. Available: <https://rekayasa.nusaputra.ac.id/article/view/86>
- [3] A. J. Putri, A. S. Syafira, M. E. Purbaya, and D. Purnomo, "Analisis Sentimen E-Commerce Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine," *J. TRINISTIK J. Tek. Ind. Bisnis Digit. dan Tek. Logistik*, vol. 1, no. 1, pp. 16–21, 2022, doi: 10.20895/trinistik.v1i1.447.
- [4] F. A. Khatami, B. Irawan, and C. Setianingsih, "Analisis Sentimen Terhadap Review Aplikasi Layanan E-Commerce Menggunakan Metode Convolutional Neural Network," *e-Proceeding Eng.*, vol. 7, no. 2, pp. 4559–4566, 2020.
- [5] A. Kurniadi, "Penerapan konsep e-bisnis pada perusahaan blibli.com," pp. 1–6, 2019.
- [6] A. Aryatama, "Analisis E-Commerce Blibli," *J. Fak. Komput.*, pp. 1–6, 2021.
- [7] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode *Naïve Bayes Classifier* Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [8] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan *Naive Bayes* Dan Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 296, 2022, doi: 10.30865/json.v4i2.5381.
- [9] D. S. Arum, S. Butsianto, R. Astuti, and U. Pelita Bangsa, "Analisis Sentimen Masyarakat Indonesia Terhadap Sea Games 2023 Di Twitter Dengan Metode *Naïve Bayes*," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 7, no. 3, pp. 728–738, 2023, doi: 10.52362/jisamar.v7i3.1150.
- [10] Riswandi, "Transaksi On-Line (E-Commerce): Peluang dan Tantangan Dalam Perspektif Ekonomi Islam," *Angew. Chemie Int. Ed. 6(11)*, 951–952., vol. 13, no. April, pp. 15–38, 2019.

- [11] Biro Pusat Statistik, “eCommerce 2022/2023 01,” pp. 1–144, 2023, [Online]. Available: <https://www.bps.go.id/>
- [12] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [13] K. Melina Elistiana, B. Adhi Kusuma, P. Subarkah, and H. Akbar Awal Rozaq, “Improvement of Naive Bayes Algorithm in Sentiment Analysis of Shopee Application Reviews on Google Play Store,” vol. 4, no. 6, pp. 1431–1436, 2023, [Online]. Available: <https://doi.org/10.52436/1.jutif.2023.4.6.1486>
- [14] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.
- [15] N. Noviyanto, “Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia,” *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.
- [16] D. D. A. Yani, H. S. Pratiwi, and H. Muhandi, “Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, p. 257, 2019, doi: 10.26418/justin.v7i4.30930.
- [17] H. Syah and A. Witanti, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm),” *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.