

ANALISIS SENTIMEN PADA PENGGUNA APLIKASI DANA MENGUNAKAN ALGORITMA NAIVE BAYES

Ines Nursatika Kusuma, Irfan Ali*

Manajemen Informatika, STMIK IKMI Cirebon
Jalan Perjuangan No.10 B Majasem Kec.Kesambi Kota Cirebon Jawa Barat
Rekayasa Perangkat Lunak
irfanaali0.0@gmail.com

ABSTRAK

Perkembangan teknologi keuangan dalam bentuk dompet digital sangat menarik perhatian masyarakat. Penggunaan e-wallet sebagai metode pembayaran modern telah memberikan kenyamanan dalam melakukan transaksi dan mengubah cara orang melakukan transaksi di era digital. Aplikasi Dana adalah platform keuangan yang menyediakan layanan keuangan untuk memfasilitasi pengguna. Pada konteks ini sangat penting untuk memahami persepsi pengguna tentang Aplikasi Dana pada masalah yang menggambarkan ketidakpuasan pelanggan mengenai kualitas layanan serta respons terhadap umpan balik pengguna dan keamanan serta privasi pengguna berdasarkan pengalaman pengguna aplikasi Dana berdasarkan perasaan positif, netral, ataupun negatif. Penelitian ini bertujuan untuk memberikan informasi terkait dengan perasaan pengguna pada aplikasi Dana. Penelitian ini menerapkan metode KDD dalam menganalisis ulasan pengguna menggunakan web scrapping dengan algoritma Naive Bayes yaitu pendekatan klasifikasi efektif dalam menganalisis sentimen. Dengan analisis sentimen ini diharapkan dapat memberikan wawasan kepada pengembang aplikasi untuk memahami terkait kelemahan pada layanan yang diberikan. Dengan adanya sentimen, pengembang dapat mengambil solusi untuk meningkatkan layanan, memperbaiki kekurangan pada aplikasi tersebut, serta meningkat kepuasan kepada penggunanya. Penelitian ini berfokus pada pemahaman mendalam tentang persepsi dan reaksi pengguna terhadap berbagai aspek aplikasi Dana. Dihasilkan sebanyak 576 ulasan positif, ulasan netral sebanyak 205 dan ulasan negatif sebanyak 219 ulasan. Berdasarkan metode Naive Bayes menghasilkan Performance Vector didapat accuracy: 74.60% +/- 3.53% (micro average: 74.60%) dan kappa: 0.550 +/- 0.061 (micro average: 0.550)

Kata kunci : Analisis sentimen, Naive Bayes, E-wallet, Dana, Pembayaran Virtual

1. PENDAHULUAN

Perkembangan informasi teknologi telah memberikan dampak yang meningkat pada berbagai aspek kehidupan manusia, termasuk pada bidang keuangan. Salah satu hasil dari transformasi digital ini adalah munculnya aplikasi keuangan berbasis teknologi, seperti aplikasi Dana. Aplikasi ini memberikan kenyamanan bagi pengguna dalam mengelola aspek keuangan, melakukan transaksi, dan mengakses layanan keuangan secara praktis. Namun, seiring dengan popularitasnya yang meningkat, jumlah pengguna yang semakin meningkat juga menimbulkan kekhawatiran tentang kualitas layanan yang diberikan oleh Dana.

Pendekatan analisis sentimen muncul sebagai metode yang efektif untuk mengukur dan memahami persepsi dan tanggapan pengguna terhadap suatu produk atau layanan. Dalam konteks aplikasi Dana, memahami bagaimana pengguna merespons pengalaman mereka sangatlah penting dengan menggunakan metode analisis sentimen. Salah satu metode yang sering digunakan dalam analisis sentimen adalah Metode Naive Bayes. Metode Naive Bayes merupakan pendekatan klasifikasi statistik yang berakar pada teorema Bayes. Metode ini telah banyak diterapkan di berbagai bidang, termasuk analisis sentimen, karena sifatnya yang sederhana namun efektif. Dalam konteks analisis sentimen review pada aplikasi Dana, penerapan Metode Naive Bayes

merupakan alternatif menarik untuk mengidentifikasi dan mengklasifikasikan sentimen positif, negatif, atau netral yang terdapat dalam ulasan pengguna.

Dana merupakan dompet digital Indonesia yang dirancang untuk menjadikan setiap transaksi secara digital, baik online maupun offline dapat digunakan dengan cepat, praktis serta tetap terjamin dalam hal keamanannya. Dana, sebagai platform dompet digital terbuka, berperan sebagai pendukung utama bagi segala aktivitas ekonomi dan gaya hidup digital di kalangan masyarakat Indonesia. Keberadaannya memungkinkan masyarakat untuk meningkatkan tingkat produktivitas, efisiensi, dan kompetensi mereka. Pada tahun 2021, Dana mencatat pertumbuhan yang signifikan dalam hal jumlah pengguna dan rata-rata volume transaksi. Jumlah pengguna Dana mengalami peningkatan sebesar 40%, melonjak dari 50 juta pada Desember 2020 menjadi 70 juta pada paruh pertama tahun 2021. Capaian ini mencerminkan komitmen Dana dalam memastikan dan meningkatkan pengalaman transaksi pengguna melalui pengembangan teknologi secara progresif. Pertumbuhan yang pesat ini juga menunjukkan bahwa Dana memiliki fokus yang jelas terhadap inklusivitas dalam teknologi finansial, dengan manfaat yang semakin meluas. Selain berperan sebagai sarana pembayaran, Dana juga mampu memberikan kontribusi dalam interaksi sosial, kegiatan bisnis, serta fungsi menabung dan berinvestasi. Pencapaian ini

mencerminkan upaya Dana dalam mengembangkan teknologi finansial yang tidak hanya efektif secara ekonomi, tetapi juga memberikan manfaat yang lebih luas bagi masyarakat.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis adalah sentimen proses analisis data, media sosial, analisis media sosial, analisis, umpan balik pengguna yang digunakan untuk mengevaluasi sentimen pengguna terhadap suatu produk, topik, atau artikel yang diminati. dan umpan balik pengguna yang digunakan untuk mengevaluasi sentimen terhadap suatu produk, atau sebuah topik, serta artikel yang menarik. Mengidentifikasi atau mengumpulkan informasi penting.[1]

Analisis sentimen merupakan komponen integral dalam proses pengolahan data yang terfokus pada klasifikasi teks ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Signifikansi analisis sentimen bagi suatu organisasi tercermin dalam kemampuannya untuk memperoleh pemahaman mendalam terhadap sentimen yang terkandung dalam teks, mengizinkan pengenalan pola serta evaluasi tanggapan yang mungkin berdampak pada keputusan dan strategi organisasi.[2]

Analisis sentimen merupakan inovasi teknologi terbaru yang saat ini tengah dalam tahap pengembangan dan penerapan dalam konteks penelitian untuk menganalisis topik dan pendapat. Teknologi ini sering digunakan untuk mengevaluasi dan menganalisis tingkat kepuasan pelanggan atau pengguna terhadap suatu produk atau kebijakan tertentu.[3]

2.2. Data Mining

Data mining merupakan teknologi untuk melakukan ekstraksi informasi dari sebuah data yang dapat dilakukan untuk berbagai tujuan, seperti: mengetahui perilaku pelanggan sesuai dengan kebiasaan konsumsi dan permintaan di masa depan, dan juga dapat memahami pelanggan bagaimana menggunakan produk dan layanan dalam mengembangkan aplikasi berdasarkan analisis para penggunanya seperti pada jejaring sosial membuat model perkiraan untuk pasar atau industri tertentu.[4]

Data mining adalah suatu teknologi yang digunakan untuk mengekstraksi informasi dari suatu kumpulan data, dan aplikasinya dapat beragam, antara lain untuk memahami perilaku konsumen sesuai dengan kebiasaan konsumsi dan permintaan di masa depan. Selain itu, teknologi ini juga bermanfaat untuk memahami bagaimana pelanggan menggunakan produk dan layanan, yang nantinya dapat menjadi dasar untuk pengembangan aplikasi melalui analisis pengguna, seperti yang terjadi pada jejaring sosial. Dengan menggunakan data mining, mungkin juga dibuat model perkiraan untuk pasar atau industri tertentu.[5]

Data mining merupakan suatu proses penggalian data dan informasi yang berskala besar dari database, yang sebelumnya tidak diketahui namun memiliki potensi pemahaman dan kegunaan yang signifikan. Proses ini sangat penting untuk pengambilan keputusan bisnis yang krusial. Data mining mencerminkan kumpulan teknik yang digunakan dengan maksud menemukan pola yang tidak teridentifikasi sebelumnya dalam data yang telah terkumpul. Dengan adanya data mining, pengguna memiliki kemampuan untuk menemukan pengetahuan yang tersembunyi dalam basis data yang tidak dapat diketahui dengan metode biasa. Dengan otomatisasi dalam pencarian informasi yang berharga dalam penyimpanan data berskala besar, data mining membentuk suatu proses yang dapat dibagi menjadi beberapa tahapan.[6]

2.3. Text Mining

Text Mining merupakan suatu proses ekstraksi informasi dari sumber data untuk tujuan analisis, di mana informasi tersebut dikelompokkan berdasarkan kata-kata untuk mengungkapkan hubungan dengan data sumber lainnya. Tugas utama dalam bidang Text Mining melibatkan dua aspek penting, yaitu pengkategorian teks (text categorization) dan pengelompokan teks (text clustering). Setelah proses pengkategorian dan pengelompokan teks dilakukan, langkah-langkah selanjutnya mencakup serangkaian prosedur seperti tokenisasi, di mana kalimat diuraikan menjadi kata-kata; kapitalisasi, yang mengubah huruf besar menjadi huruf kecil; stemming, yang mereduksi kata-kata menjadi bentuk dasarnya; dan penggunaan teknik perhitungan serta pengelompokan kata (filtering). Penerapan Text Mining mencakup berbagai aspek seperti analisis sentimen, pemahaman preferensi pelanggan, penyaringan spam, dan berbagai keperluan lainnya.[7]

Text Mining merupakan penerapan prinsip-prinsip dan teknik data mining yang difokuskan pada penemuan pola-pola yang terdapat dalam teks. Proses analisis teks ini dilakukan dengan maksud untuk mengidentifikasi informasi yang memiliki nilai dan relevansi khusus. Dalam konteks lain yang berkaitan dengan text mining, didefinisikan bahwa ini merupakan kegiatan penambangan data yang dilakukan pada teks, dimana sumber data umumnya berasal dari dokumen. Tujuan dari kegiatan ini adalah untuk mengidentifikasi kata-kata yang mampu mencerminkan substansi atau inti dari dokumen tersebut.[8]

2.4. Naive Bayes

Naive Bayes adalah suatu metode klasifikasi probabilistik yang sederhana, yang melakukan perhitungan probabilitas dengan merangkum frekuensi dan kombinasi nilai dari dataset yang diberikan. Pendekatan ini didasarkan pada asumsi sederhana bahwa nilai atribut bersifat kondisional saling bebas jika diberikan nilai output. Dengan kata lain,

probabilitas bersama dari mengamati nilai output dihitung sebagai hasil perkalian probabilitas individu atribut. Metode klasifikasi Naive Bayes menonjol karena rumusnya yang sederhana dan kemudahan aplikasinya, sambil tetap mencapai tingkat akurasi yang cukup tinggi jika dibandingkan dengan metode-metode lainnya. Naive Bayes termasuk dalam kategori algoritma klasifikasi dan menerapkan pendekatan probabilitas dan statistik. Diakui sebagai temuan ilmuwan Inggris, Thomas Bayes, metode ini bertujuan untuk memprediksi peluang di masa depan berdasarkan pengalaman masa sebelumnya, sehingga sering disebut sebagai Teorema Bayes. Teorema ini kemudian diterapkan secara "naive" dengan asumsi bahwa kondisi independensi antar atribut. Dalam konteks klasifikasi Naive Bayes, diasumsikan bahwa keberadaan atau tidaknya suatu ciri dari suatu kelas tidak memiliki kaitan dengan ciri dari kelas lainnya.[9]

Naive Bayes merupakan metode pengklasifikasian probabilistik yang sederhana, yang melakukan perhitungan probabilitas dengan menggabungkan frekuensi dan kombinasi nilai-nilai dari dataset yang diberikan. Algoritma ini menerapkan teorema Bayes dengan asumsi bahwa semua atribut

bersifat independen atau tidak saling tergantung, sambil mempertimbangkan nilai dari variabel kelas. Dalam konteks metode Naive Bayes, probabilitas yang terlibat diperoleh melalui perhitungan frekuensi kemunculan atribut-atribut tertentu dalam data pelatihan. Walaupun asumsi tentang independensi semua atribut dalam Naive Bayes sering kali tidak sepenuhnya terpenuhi dalam situasi dunia nyata, namun algoritma ini sering memberikan hasil yang baik dalam implementasinya. Terutama, hal ini berlaku ketika data pelatihan memiliki ukuran yang cukup besar dan kompleksitas model tidak terlalu tinggi. Naive Bayes umumnya diterapkan dalam berbagai bidang, seperti analisis teks, klasifikasi dokumen, analisis sentimen, dan sebagainya. Kecepatan komputasinya yang relatif tinggi dan hasil yang memuaskan dalam skenario tertentu menjadikannya sebagai pilihan yang populer dalam pengklasifikasian data.[10]

Berikut Merupakan data yang dihasilkan melalui Web Scrapping kemudian dilakukan Proses pelabelan Berdasarkan hasil ulasan:

Tabel 1. Data yang dihasilkan melalui Web Scrapping

Username	Rating	Tanggal	Content
Ali Hamzah	3	12/22/2023 7:49:00 AM	akuberi 3 dulu ,banyak peguna yg kecewa, karena begitu rumit untuk datar ke Dana premiumnya. coba
David samuel	5	12/22/2023 7:48:00 AM	Sangat bagus dan gampang di oprasikan
Siti Nurhaya	5	12/22/2023 7:45:00 AM	sangat membantu
Muhamad aziz Barokah	2	12/22/2023 7:41:00 AM	Aplikasi goblog ngaco mulu
Hana Cantik	3	12/22/2023 7:38:00 AM	Dana makin parah untuk liat saldo saja susah bnget

Berdasarkan data diatas diambil melalui teknik scrapping menggunakan Google Colab yang diambil berdasarkan ulasan pengguna Aplikasi Dana sebanyak 1000 ulasan pada Google Play Store. Berikutnya proses Pelabelan berdasarkan sentimen Positif, Netral, dan Negatif.

2.5. Contoh Persamaan Matematika

Recall adalah proporsi dari prediksi benar positif terhadap keseluruhan data yang sebenarnya positif. Rumus Recall dapat diwakili sebagai

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

Sementara itu, Precision merupakan rasio dari prediksi benar positif terhadap keseluruhan hasil yang diprediksi sebagai positif. Rumus Precision dapat dinyatakan sebagai

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

3. METODE PENELITIAN

Sumber data yang dimanfaatkan dalam penelitian ini diperoleh dari evaluasi dan komentar yang tercantum di Google Play Store. Pengumpulan data dilakukan dengan menggunakan teknik web scraping untuk memastikan penggunaan data yang paling mutakhir. Data yang dikumpulkan bersumber dari halaman Google Play Store yang terkait dengan

layanan aplikasi Dana, yang dapat diakses melalui tautan:

<https://play.google.com/store/apps/details?id=id.dana>.

Proses pengambilan data ini dilakukan pada bulan Desember 2023, di mana sejumlah 1000 ulasan diambil untuk analisis.

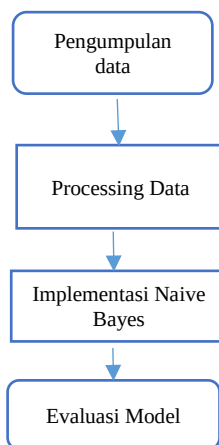
Setelah proses pengambilan data, dilakukan serangkaian tahapan preprocessing guna menyiapkan data tersebut untuk analisis sentimen menggunakan metode Naive Bayes pada ulasan Google Play Store.

Pada penelitian menggunakan teknik web scraping menggunakan Google Colab. Teknik web scraping memungkinkan untuk mengambil data dari Google play store berdasarkan kata kunci atau filter tertentu, dalam hal ini kata kunci yang digunakan adalah id.dana. Proses pengumpulan data Menggunakan data ulasan yang terbaru yang dilakukan pada bulan Desember 2023 sebanyak 1000 ulasan pengguna dana. Setelah pengumpulan data, dilakukan menganalisis dilakukan klasifikasi pada data ulasan untuk menentukan ulasan yang bersifat positif, netral, dan negatif. Proses klasifikasi data yang dikumpulkan menggunakan metode Naive Bayes.

Berikut alur dalam pengambilan data:

- 1) Google play store
- 2) Scrapping data
- 3) Format file Csv.

Pada penelitian ini dilakukan proses menggunakan klasifikasi sentimen yang sudah didapat pada ulasan aplikasi dana dan diimplementasi menggunakan Algoritma Naive Bayes sehingga diperoleh hasil akurasi sebuah sentimen. Penelitian ini memanfaatkan Metode KDD adalah pendekatan analitis dan terprogram untuk analitis data yang memodelkan data dari database untuk mengekstraksi pengetahuan yang relevan dan berguna. Pendekatan analitis dan terprogram untuk penambahan data yang memodelkan data dari database untuk mengekstraksi pengetahuan yang relevan dan berguna. Ekstraksi informasi dari kumpulan data yang sangat besar adalah tujuan utama teknik KDD. Pola dari data yang diolah menggunakan sejumlah algoritma. Ekstraksi dari kumpulan data besar adalah tujuan utama teknik KDD yang merupakan pengetahuan menggunakan teknik data mining. Studi sistematis studi, eksplorasi, dan pemodelan sumber data penting data dikenal sebagai KDD. Berikut tahapan penelitian yang dilakukan:



Gambar 1. Tahapan Penelitian yang dilakukan

Berikut tahapan penelitian yang dilakukan:

- Pengumpulan Data:**
Pengambilan sampel dari ulasan pengguna aplikasi Dana. Sebelumnya harus memiliki kumpulan data yang mencakup berbagai jenis ulasan, termasuk ulasan positif, netral, dan negatif.
- Pengumpulan dan Preprocessing Data:**
Mengumpulkan ulasan pengguna Aplikasi Dana serta melakukan preprocessing data seperti pembersihan teks.
- Implementasi Algoritma Naive Bayes:**
Mengimplementasikan algoritma Naive Bayes dalam klasifikasi sentimen ulasan, dengan memanfaatkan data training yang telah disiapkan.
- Evaluasi Model:**
Melakukan penilaian kinerja model Naive Bayes dengan menggunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score, bertujuan untuk mengevaluasi sejauh mana kemampuan

model dalam melakukan klasifikasi sentimen ulasan secara akurat.

4. HASIL DAN PEMBAHASAN

4.1. Mengetahui pengambilan dataset sentimen analisis pengguna aplikasi Dana pada Google play store.

Proses pengambilan Data:

```

[3] 1
    2
    3 from google_play_scraper import Sort, reviews
    4
    5 result, continuation_token = reviews(
    6     'id.dana',
    7     lang='id',
    8     country='id',
    9     sort=Sort.NEWEST,
   10     count=1000,
   11     filter_score_with=None
   12 )
  
```

Gambar 2. Scraping Data

Dataset yang diambil kemudian disimpan dan diubah menjadi file csv dengan data yang sesuai pada ulasan aplikasi Dana. Berikut merupakan proses perubahan dataset ke dalam format file csv.

Berikut file yang dihasilkan:

```

[10] 1 my_df.to_csv("scrapped_data.csv", index = False)
  
```

Gambar 3. Format File CSV

Pada Gambar diatas setelah dilakukannya scrapping data yang akan menampilkan format File dan disimpan dalam format csv, pada file tersebut menampilkan ulasan komentar para pengguna layanan pada Aplikasi Dana.

Berikut merupakan tabel setelah dilakukan pelabelan berdasarkan sentimen:

Tabel 2. Proses Pelabelan

SENTIMEN	CONTENT
Netral	akuberi 3 dulu ,banyak peguna yg kecewa, karena begitu rumit untuk datar ke Dana premiumnya. coba otomatis gak usa terlalu rumit.
Positif	Sangat bagus dan gampang di oprasikan
Positif	sangat membantu
Positif	mantap
Negatif	Aplikasi goblog ngaco mulu
Negatif	Dana makin parah untuk liat saldo saja susah bnget

Penelitian ini dilaksanakan dengan melibatkan data ulasan, di mana proses pengklasifikasian data dilakukan secara manual melalui pelebelan data, yang melibatkan pengelompokan data berdasarkan kategori sentimen yang telah ditentukan sebelumnya. Pengelompokan ini memiliki tujuan untuk mengevaluasi kualitas layanan yang diterima oleh pengguna aplikasi Dana.

Berikut merupakan hasil jumlah sentimen:

Tabel 3. Jumlah Sentimen

Sentimen	Jumlah
Positif	576
Netral	219
Negatif	205

Berikut jumlah yang dihasilkan setelah melakukan pelabelan berdasarkan kelas sentimen positif,netral,dan negatif pada pengguna aplikasi Dana.

4.2. Implementasi Dataset analisa sentimen positif,netral,dan negatif pada pengguna aplikasi Dana.

a. Ulasan Positif

Data ulasan yang menggambarkan ulasan positif telah diperoleh melalui proses pelabelan manual. Dalam hal ulasan yang mencerminkan sentimen positif, pelabelan dilakukan secara terperinci pada identifikasi ulasan positif dari pengguna Dana yang paling sering memberikan ulasan. Sebanyak 576 ulasan positif berhasil diidentifikasi berdasarkan proses pelabelan pada sejumlah pengguna tertentu.



Gambar 4. Visualisasi Wordcloud ulasan Positif

Visualisasi Wordcloud memungkinkan pengamatan yang lebih terperinci terhadap kata-kata positif yang sering digunakan oleh pengguna saat memberikan ulasan. Ukuran kata dalam wordcloud mencerminkan frekuensinya, sehingga semakin besar ukuran kata tersebut, semakin tinggi frekuensi pengguna menggunakannya sebagai ungkapan positif dalam penilaian mereka.

b. Ulasan Netral

Data ulasan yang menggambarkan ulasan netral diperoleh melalui proses pelabelan manual. Dalam konteks ulasan netral, pelabelan dilakukan secara terperinci pada pengidentifikasian ulasan netral dari pengguna Dana yang paling sering memberikan ulasan. Sejumlah 205 ulasan netral berhasil diidentifikasi berdasarkan hasil pelabelan pada sejumlah pengguna tertentu.



Gambar 5. Visualisasi Wordcloud ulasan Netral

Visualisasi wordcloud memungkinkan pengamatan yang lebih rinci terhadap kata-kata netral

yang sering digunakan oleh pengguna dalam memberikan ulasan. Ukuran kata dalam wordcloud mencerminkan frekuensinya, sehingga semakin besar ukuran kata tersebut, semakin tinggi frekuensi pengguna menggunakan kata tersebut sebagai subjek pembicaraan dengan penilaian netral dalam ulasan mereka.

c. Ulasan Negatif

Pada ulasan yang mencerminkan sentimen negatif, pelabelan dilakukan secara teliti hingga berhasil mengidentifikasi ulasan negatif dari pengguna Dana yang sering memberikan ulasan. Berdasarkan hasil pelabelan, terdapat sebanyak 219 ulasan negatif yang berhasil diidentifikasi pada sejumlah pengguna Dana tertentu.

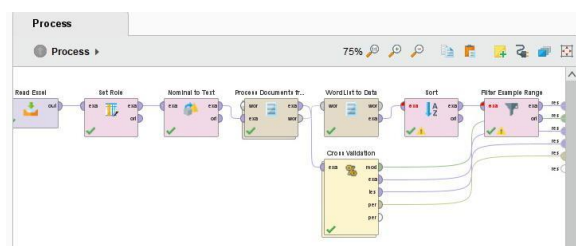


Gambar 6. Visualisasi Wordcloud ulasan Negatif

Visualisasi wordcloud memungkinkan pengamatan yang lebih terinci terhadap kata-kata negatif yang sering digunakan oleh pengguna saat memberikan ulasan. Ukuran kata dalam wordcloud mencerminkan frekuensinya, sehingga semakin besar ukuran kata tersebut, semakin tinggi frekuensi pengguna menggunakannya sebagai ungkapan negatif dalam penilaian mereka.

4.3. Evaluasi Dataset sentimen analisis dalam menerapkan metode Naive bayes

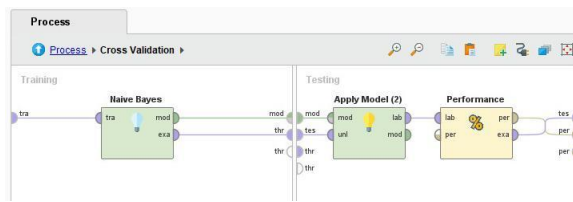
Berikut beberapa operator yang digunakan:



Gambar 7. Model Klasifikasi Naive Bayes

Pada Gambar diatas menggunakan beberapa operator yang digunakan diantaranya: Read Excel, Set Role, Nominal to Text, Process Documents from Data, Word list to Data, Sort, Filter Exemple Range, Cross Validation. Proses Klasifikasi menggunakan Algoritma Naive Bayes.

Berikut merupakan tampilan operator pada Cross Validation.



Gambar 8. Model Klasifikasi Naive Bayes 1

Pada Cross validation terdapat beberapa operator yang digunakan yaitu Naive Bayes, Apply Model, dan Performance.

a. Naive Bayes:

Metode pengklasifikasi yang paling banyak digunakan dan dikenal dengan tingkat keakuratan yang tinggi adalah Naive Bayes. Banyak penelitian sebelumnya telah melibatkan penggunaan algoritma ini dalam konteks pengklasifikasi.

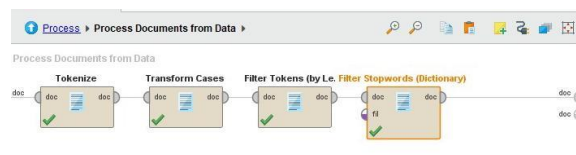
b. Apply Model:

Operator ini menerapkan model pada Contoh Set. Biasanya tujuan operator ini adalah untuk memprediksi pada data yang tidak terlihat untuk mentransformasikan data dengan menerapkan model preprocessing.

c. Performance:

Performance adalah teknik yang digunakan untuk memprediksi keanggotaan kelompok untuk contoh data.

Berikut tampilan pada Operator Documents From Data



Gambar 9. Documents From Data

Pada operator Process Documents From Data menampilkan beberapa operator diantaranya: Tokenize, Transform cases, Filter Token by length, dan Filter Stopwords (Dictionary).

a. Tokenize:

Operator ini membagi teks dokumen menjadi serangkaian token. Ada beberapa opsi yang digunakan untuk menentukan titik pemisahan. Bisa menggunakan semua karakter non-huruf, apa pengaturan defaultnya.

b. Transform Cases:

Operator ini mentransformasikan seluruh karakter dalam dokumen menjadi huruf kecil atau huruf besar tertentu akan dihapus.

c. Filter Token By length:

Operator ini melakukan filtrasi pada token berdasarkan panjang kata, dengan mengeliminasi kata-kata yang memiliki jumlah karakter tertentu dalam upaya untuk mengambil kata-kata yang dianggap penting dari hasil tokenisasi.

d. Stopword:

Pada Operator ini memfilter stopwords bahasa Inggris dari dokumen dengan menghapus setiap token yang sama dengan stopwords dari daftar stopwords bawaan.

Berikut yang dihasilkan Performance vector Accuracy dan Kappa:

accuracy: 74.60% +/- 3.53% (micro average: 74.60%)

	true netral	true positif	true negatif	class precision
pred. netral	31	23	64	51.12%
pred. positif	78	525	11	85.50%
pred. negatif	56	28	130	62.50%
class recall	41.55%	91.15%	63.41%	

Gambar 10. Performance Vector Accuracy

berdasarkan Gambar dihasilkan accuracy: 74.60% +/- 3.53% (micro average: 74.60%. class precision positif 51.12%, netral 51.12%, dan negatif 62.50. sedangkan berdasarkan class recall positif 91.15%, netral 41.55%, dan negatif 63.41%.

kappa: 0.550 +/- 0.061 (micro average: 0.550)

	true netral	true positif	true negatif	class precision
pred. netral	31	23	64	51.12%
pred. positif	78	525	11	85.50%
pred. negatif	56	28	130	62.50%
class recall	41.55%	91.15%	63.41%	

Gambar 11. Performance Vector Kappa

berdasarkan Gambar dihasilkan kappa: 0.550 +/- 0.061 (micro average: 0.550) class precision positif 51.12%, netral 51.12%, dan negatif 62.50. sedangkan berdasarkan class recall positif 91.15%, netral 41.55%, dan negatif 63.41%.

Berikut kata yang paling banyak diulas:

Row No.	word	in documents	total	in class (pos.)	in class (neg.)	in class (net.)
1	dana	178	202	67	29	129
2	saya	116	200	81	55	129
3	sangat	178	185	5	965	15
4	biasa	124	160	60	81	31
5	bagus	105	160	1	159	1
6	tidak	105	111	25	28	40
7	mantap	93	100	41	4	61
8	membantu	84	94	1	33	9
9	saldo	84	85	2	79	4
10	ada	62	85	24	0	61

Gambar 12. Tampilan kata yang paling banyak digunakan

Pada Gambar diatas merupakan kata yang paling banyak ditulis oleh para pengguna aplikasi Dana. Kata Dana ada sebanyak 178, kata Saya ada 116, kata Sangat ada sebanyak 178, kata Biasa ada sebanyak 124, kata Bagus ada sebanyak 105, kata Tidak ada sebanyak 83, kata Mantap ada sebanyak 93, kata Membantu ada sebanyak 84, kata Saldo ada sebanyak 62.

Tampilan yang dihasilkan pada Simple Distribution:

Class	netral	pos.	neg.
Class netral	0.5129	0.5129	0.5129
Class positif	0.5129	0.5129	0.5129
Class negatif	0.5129	0.5129	0.5129

Gambar 13. Simple Distribution

Pada ilustrasi di atas, tergambar hasil dari proses klasifikasi dengan menggunakan metode Naive Bayes yang menghasilkan tiga kelas yang ada dalam distribusi sederhana, yaitu: pada Simple Distribution diperoleh class netral 0.218 dan 1499 distributions, class positif 0.573 dan 1499 distributions, class negatif 0.024 dan 1499 distributions.

5. KESIMPULAN DAN SARAN

Proses pengambilan data menggunakan Scrapping pada Google Colab yang diambil dari ulasan pengguna aplikasi Dana. Pada Persepsi pengguna aplikasi Dana pada website Google play berdasarkan ulasan pada bulan Desember 2023, berdasarkan klasifikasi sentimen positif sebanyak 579 ulasan, sentimen netral sebanyak 219 ulasan, sentimen negatif sebanyak 205 ulasan. Sedangkan hasil yang diperoleh dari penggunaan metode Naive Bayes dalam mengklasifikasi data ulasan pengguna aplikasi Dana dibagi menjadi kelas positif, netral, dan negatif. Diperoleh hasil klasifikasi sentimen dengan tingkat akurasi sebesar Performance Vector yang didapat accuracy: 74.60% +/- 3.53% (micro average: 74.60%, class precision positif 51.12%, netral 51.12%, dan negatif 62.50, sedangkan berdasarkan class recall positif 91.15%, netral, dan juga hasil kapabilitas diperoleh 0.550 +/- 0.061 (micro average: 0.550) class precision positif 51.12%, netral 51.12%, dan negatif 62.50, sedangkan berdasarkan class recall positif 91.15%, netral 41.55%, dan negatif 63.41%. Sedangkan kata yang paling banyak ditulis oleh para pengguna aplikasi Dana. Kata Dana ada sebanyak 178, kata Saya ada 116, kata Sangat ada sebanyak 178, kata Biasa ada sebanyak 124, kata Bagus ada sebanyak 153, kata Aplikasi sebanyak 105, kata Tidak ada sebanyak 83, kata Mantap ada sebanyak 93, kata Membantu ada sebanyak 84, kata Saldo ada sebanyak 62. Dan hasil dari Simple Distribution dari proses klasifikasi dengan menggunakan metode Naive Bayes yang menghasilkan tiga kelas yang ada dalam distribusi sederhana, yaitu: pada Simple Distribution diperoleh class netral 0.218 dan 1499 distributions, class positif 0.573 dan 1499 distributions, class negatif 0.024 dan 1499 distributions. Dari hasil analisis dan simpulan yang diperoleh, dapat disampaikan beberapa rekomendasi. Dalam konteks penelitian ini, terdapat pembatasan pada penggunaan data ulasan yang terbatas pada bahasa Indonesia, sehingga disarankan untuk memperluas cakupan pada penelitian mendatang dengan melibatkan ulasan dalam berbagai bahasa. Selain itu, bagi peneliti selanjutnya, disarankan untuk menjelajahi pendekatan machine learning lain sebagai pembandingan dalam mengevaluasi kinerja algoritma Naive Bayes dalam mengklasifikasi ulasan aplikasi Dana atau pada platform lain selain Google Play.

DAFTAR PUSTAKA

- [1] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- [2] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Appl. Sci.*, vol. 13, no. 7, 2023, doi: 10.3390/app13074550.
- [3] Kokom Komariyah, Rahaditya Dasuki, Dias Bayu Saputra, Saeful Anwar, and Gifthera Dwilestari, "Klasifikasi Stok Barang Menggunakan Algoritma Naive Bayes Pada Pt. Dharma Electrindo Manufacturing," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 2, pp. 35–41, 2020, doi: 10.32485/kopertip.v4i2.117.
- [4] C. Liang *et al.*, "Intrusion detection system for the internet of things based on blockchain and multi-agent systems," *Electron.*, vol. 9, no. 7, pp. 1–27, 2020, doi: 10.3390/electronics9071120.
- [5] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [6] Rayuwati, Husna Gemasih, and Irma Nizar, "IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID," *Jurnal Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [7] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," ... *Teknol. Inf. dan ...*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] Nitha Kumala Dewi, "Identifikasi Berita Hoax dengan Menerapkan Algoritma Text Mining," *J. Informatics, Electr. Electron. Eng.*, vol. 2, no. 3, pp. 65–74, 2023, doi: 10.47065/jieeee.v2i3.888.
- [9] L. Siburian, "RESOLUSI: Rekayasa Teknik Informatika dan Informasi Data Mining Memprediksi Kebutuhan Vaksin Imunisasi dengan Menggunakan Metode Naive Bayes (Studi kasus UPT Puskesmas Teladan)," *Data Min. Memprediksi Kebutuhan Vaksin Imunisasi dengan Menggunakan Metod. Naive Bayes (Studi kasus UPT Puskesmas Teladan)*, vol. 1, no. 5, pp. 282–290, 2021, [Online]. Available: <https://djournals.com/resolusi>
- [10] S. A. R. Rizaldi, S. Alam, and I. Kurniawan, "Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 109–117, 2023, doi: 10.55123/storage.v2i3.2334.