JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i3.6191

PENERAPAN ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI ADAKAMI DI GOOGLE PLAY STORE

Anwar Setia Mubarokah[1], Bambang Irawan[2], Willy Prihartono[3]

[1,2,3]STMIK IKMI CIREBON; Jl. Perjuangan No 10 B Maja asem, Kesambi, Cirebon, Jawa Barat

Keywords:
3-5 keyword;
Algorithm a;
B algorithms;
Complexity.

Abstrak. Aplikasi pinjaman online semakin populer di Indonesia, termasuk aplikasi AdaKami yang menjadi salah satu platform keuangan digital terkemuka. Namun, ulasan pengguna di *Google PlayStore* menunjukkan beragam sentimen. Bagi calon pengguna, ulasan dan rating aplikasi dapat memberikan gambaran yang lebih jelas tentang pengalaman penggunaan aplikasi tersebut. Tujuan dari penelitian ini adalah untuk menganalisis sentimen ulasan pengguna terhadap aplikasi AdaKami dengan menggunakan algoritma *Naive Bayes*. Algoritma ini dipilih karena kemampuannya yang baik dalam mengklasifikasikan teks. Dataset yang digunakan terdiri dari ulasan pengguna dalam bahasa Indonesia yang diambil dari *Google PlayStore*. Proses penelitian mencakup beberapa tahapan, yaitu pengumpulan data, pemilihan data, pra-pemrosesan data (termasuk tokenisasi, penghilangan *stopword*, dan *stemming*), pembobotan teks menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), dan klasifikasi sentimen menggunakan algoritma *Naive Bayes*.Hasil evaluasi dari algoritma *Naïve bayes* dalam mengklasifikasikan sentimen ulasan pengguna AdaKami dengan nilai akurasi 85%, presisi 92%, recall 85%, dan f1-score 89%.Hasil penelitian diharapkan dapat memberikan penilaian yang bermanfaat bagi calon pengguna mengenai persepsi umum terhadap aplikasi AdaKami, serta memberikan wawasan bagi pengembang aplikasi.

Abstract. *Online lending apps are increasingly popular in Indonesia, including the AdaKami app which is one of the leading digital finance platforms. However, user reviews on Google PlayStore show mixed sentiments. For potential users, app reviews and ratings can provide a clearer picture of the app usage experience. The purpose of this study is to analyze the sentiment of user reviews of the AdaKami application using the Naive Bayes algorithm. This algorithm was chosen because of its good ability to classify text. The dataset used consists of user reviews in Indonesian taken from Google PlayStore. The research process includes several stages, namely data collection, data selection, data pre-processing (including tokenization, stopword removal, and stemming), text weighting using the Term Frequency-Inverse Document Frequency (TF-IDF) method, and sentiment classification using the Naive Bayes algorithm. The evaluation results of the Naïve Bayes algorithm in classifying the sentiment of AdaKami user reviews with an accuracy value of 85%, precision of 92%, recall of 85%, and f1-score of 89%.The results of the study are expected to provide a useful assessment for potential users regarding the general perception of the AdaKami application, as well as providing insight for application developers.*

1. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah banyak aspek kehidupan, termasuk sektor keuangan, dengan munculnya berbagai layanan keuangan digital, seperti aplikasi pinjaman online. Salah satu aplikasi pinjaman online yang populer di Indonesia adalah AdaKami, yang menawarkan kemudahan akses pinjaman tanpa prosedur yang rumit, berbeda dengan lembaga keuangan tradisional. Namun, meskipun aplikasi ini memberikan banyak manfaat, ulasan pengguna di Google Play Store menunjukkan beragam sentimen, mulai dari kepuasan terhadap kemudahan layanan hingga keluhan mengenai bunga tinggi dan kebijakan penagihan. Salah satu cara untuk mengetahui pendapat dan perasaan pengguna adalah dengan menganalisis ulasan pengguna yang tersedia pada Google Play Store [1].Dalam era digital saat ini, Google Play Store telah menjadi salah satu platform terkemuka bagi pengguna Android untuk mengakses dan mengunduh berbagai aplikasi. [2] Seiring banyaknya penyedia pinjaman online, masyarakat akan lebih banyak lagi membicarakan aplikasi pinjaman online. Banyak berbagai tempat Masyarakat bisa menyuarakan opini atau pendapatnya, mulai dari mulut ke mulut sampai melalui sosial media. Pendapat yang disampaikan dari masyarakat terhadap pinjaman online (pinjol) berupa positif dan negatifnya dari penyedia jasa pinjamanonlinetersebut [3].

Masalah ini menjadi penting karena persepsi pengguna dapat mempengaruhi reputasi aplikasi dan daya saingnya di pasar.Bagi calon pengguna, ulasan dan rating aplikasi menjadi referensi penting untuk memahami pengalaman orang lain. Namun, dengan jumlah ulasan yang banyak, terutama dalam bahasa Indonesia, pengolahan secara manual menjadi tidak efisien. Oleh karena itu, dibutuhkan pendekatan berbasis teknologi, seperti analisis sentimen menggunakan algoritma untuk memberikan pemahaman yang lebih terstruktur dan efisien [4].Penelitian sebelumnya telah menunjukkan efektivitas algoritma Naive Bayes dalam mengklasifikasikan sentimen pada teks bahasa Indonesia, terutama dalam aplikasi berbasis ulasan pengguna [5].

Namun, masih sedikit penelitian yang secara khusus menganalisis sentimen ulasan terhadap aplikasi AdaKami menggunakan algoritma Naive Bayes. Algoritma Naive Bayes dipilih karena efisiensi dan implementasinya yang mudah.[6] Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan menganalisis sentimen ulasan pengguna terhadap aplikasi AdaKami di Google Play Store menggunakan metode Naive Bayes. Penerapan metode Naïve Bayes dalam analisis sentimen ulasan aplikasi di Google Play Store telah terbukti efektif dalam mengklasifikasikan sentimen pengguna.[7]

Penelitian ini diharapkan dapat memberikan kontribusi praktis bagi pengembang aplikasi dan pihak terkait dalam meningkatkan kualitas layanan berdasarkan masukan pengguna. Tujuan dari penelitian ini adalah untuk mengklasifikasikan dan mengetahui sentimen pengguna aplikasi Satu Sehat menggunakan algoritma Naïve Bayes Classifier[8] Selain itu, secara akademis, penelitian ini memperkaya literatur tentang analisis sentimen dalam bahasa Indonesia, khususnya dalam konteks aplikasi pinjaman online.

2. TINJAUAN PUSTAKA

Analisis sentimen adalah proses mengidentifikasi dan mengkategorikan emosi (positif, negatif, atau netral) yang terkandung dalam teks dengan menggunakan teknik analisis teks. Analisis sentimen sering dipakai untuk mengungkap opini atau pandangan pelanggan, pengguna, atau masyarakat terhadap suatu produk, layanan, atau gagasan.[1]

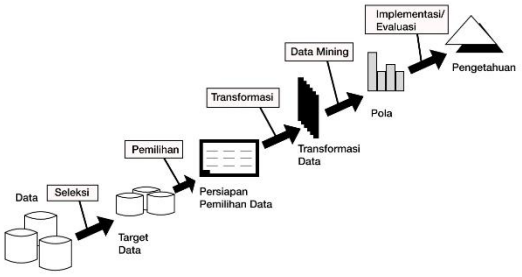
Pemberian rating aplikasi di Google Play Store diikuti dengan ulasan dari para pengguna terhadap aplikasi tersebut. Ulasan tersebut mengandung opini dari para pengguna mengenai aplikasi tersebut dan calon pengguna melihat ulasan dari sebuah aplikasi sebagai pertimbangan sebelum memutuskan untuk menggunakan aplikasi tersebut.[9]

Multinomial Naive Bayes merupakan suatu pendekatan yang bekerja dengan menghitung frekuensi dari masing-masing term dalam dokumen . Kategori dokumen pada klasifikasi Multinomial Naïve Bayes tidak hanya berdasarkan kata-kata yang muncul di dalamnya, tetapi juga mempertimbangkan jumlah kemunculan dari kata-kata tersebut. Sehingga dalam konteks Multinomial Naive Bayes ini peran tokenisasi sangat penting[10]

3. METODE PENELITIAN

proses pemilihan data yang relevan dan dapat dilakukan analisis dari data operasional [11].Proses ini memilih dan menyaring data yang relevan dan sesuai dengan tujuan penelitian atau 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. Proses KDD bertujuan untuk mengubah data mentah menjadi informasi yang dapat digunakan untuk pengambilan keputusan [12].

pendekatan analitis dan terprogram untuk penambangan 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 minin . Studi sistematis studi, eksplorasi, dan, pemodelan sumber data penting data dikenal sebagai KDD.[13]. Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:



Gambar 1. Penelitian Metode KDD

2.1. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini dikumpulkan dari *Google Play Store*, khususnya ulasan pengguna terhadap aplikasi AdaKami. Pengumpulan data dilakukan dengan menggunakan metode web *scraping* untuk mengambil ulasan pengguna dari halaman aplikasi AdaKami di *Google Play Store* sebanyak 1000 data. Web scraping adalah proses pengambilan data dari situs web secara otomatis tanpa harus menyalinnya secara manual [14].*Tools* yang digunakan adalah *Python* dengan *library google-play-scraper.*

Data yang dikumpulkan meliputi teks ulasan, rating (nilai bintang yang diberikan pengguna), dan tanggal ulasan. Selain itu, informasi seperti nama pengguna dan versi aplikasi yang digunakan juga dapat dikumpulkan sebagai data tambahan, meskipun fokus utama adalah teks ulasan dan rating.

Proses pengumpulan dimulai dengan mengakses halaman aplikasi AdaKami di *Google Play Store*. Setelah itu, web scraping dilakukan untuk mengambil data ulasan dengan menggunakan skrip *Python*. Data ulasan yang terkumpul kemudian disaring dan disimpan dalam format CSV untuk memudahkan analisis lebih lanjut.

2.2. Tahap penelitian

Tahapan penelitian ini yaitu metode Knowledge Discovery in Data (KDD) yang terdiri dari tahapan data selection, data cleaning, data transformation, data mining, pattern evaluation, dan knowledge presentation.

a. Data Selection
   Data selection adalah analisis, dengan mempertimbangkan kriteria tertentu seperti kualitas, representasi, dan kesesuaian dengan masalah yang akan diselesaikan. Proses ini bertujuan untuk memastikan bahwa data yang digunakan memiliki relevansi tinggi dan dapat memberikan hasil analisis yang akurat serta mendukung pengambilan keputusan yang efektif.

b. Preprocessing Data
   Preprocessing data adalah tahap awal dalam analisis data yang bertujuan untuk mempersiapkan data mentah agar siap digunakan dalam proses analisis lebih lanjut. Proses preprocessing data teks melibatkan beberapa tahapan, antara lain penghapusan kata non-Inggris, penghapusan tanda baca, penghapusan emotikon, tokenisasi, dan standarisasi teks[15]. Pada tahap ini, data yang dikumpulkan melalui teknik seperti web scraping akan dibersihkan dari elemen yang tidak relevan, seperti tanda baca, angka, dan karakter khusus.

c. Data Transformation
   Transformasi data yang efektif dapat membantu dalam mengurangi kompleksitas data dan meningkatkan efisiensi proses analisis.Proses mengubah data dari format atau struktur aslinya menjadi format yang lebih sesuai untuk analisis, pemrosesan, atau integrasi. Langkah ini sangat penting dalam mempersiapkan data untuk analisis lebih lanjut, karena memastikan konsistensi, akurasi, dan kompatibilitas dengan alat atau algoritma yang digunakan. Transformasi data dapat melibatkan berbagai operasi, seperti penggabungan data, normalisasi nilai, pengkodean variabel kategorikal, penskalaan fitur numerik, atau perubahan struktur data. Tujuan dari transformasi data adalah untuk membuat data lebih bermakna dan lebih mudah diinterpretasikan, sehingga memungkinkan analisis yang lebih efektif dan pengambilan keputusan yang lebih baik.

d. Data Mining
   Proses ekstraksi pola, informasi, atau pengetahuan yang berguna dari kumpulan data besar dengan menggunakan teknik statistik, algoritma, dan pembelajaran mesin. Tujuan utama dari data mining adalah untuk mengidentifikasi pola tersembunyi dalam data yang dapat digunakan untuk pengambilan keputusan yang lebih baik, prediksi, atau pemahaman lebih dalam terhadap fenomena tertentu. Proses ini mencakup beberapa tahap, seperti pembersihan data, transformasi data, penerapan algoritma analisis, dan interpretasi hasil. Data mining sering digunakan dalam berbagai bidang, termasuk bisnis, kesehatan, keuangan, dan pemasaran, untuk menemukan tren, pola perilaku, atau hubungan yang tidak terlihat secara langsung.

e. Evaluasai
   Proses untuk menilai kinerja atau efektivitas suatu model, metode, atau sistem berdasarkan hasil yang diperoleh. Dalam konteks penelitian atau analisis data, evaluasi dilakukan untuk mengukur sejauh mana model atau pendekatan yang digunakan berhasil mencapai tujuan yang telah ditetapkan, dengan menggunakan berbagai metrik atau indikator kinerja seperti akurasi, presisi, recall, atau F1-score. Evaluasi ini penting untuk

memastikan bahwa hasil yang diperoleh valid dan dapat diandalkan.

4. HASIL DAN PEMBAHASAN

4.1 PEMBAHASAN

Berdasarkan hasil penelitian yang telah dilakukan, pembahasan berikut akan difokuskan pada evaluasi performa algoritma *Naïve Bayes Classifier* dalam analisis sentimen ulasan aplikasi AdaKami. Evaluasi ini akan mencakup perhitungan berbagai metrik, yaitu akurasi *(accuracy)*, *precision*, *recall*, *F1-score* dan matriks kebingungan *(confusion matrix)* yang digunakan untuk mengukur efektivitas algoritma *Naïve Bayes Classifier*.

4.1.1 Data Selection

Tahapan ini adalah Data *Selection*.disini data di seleksi bahasa indonesia yang relevan dan hasil nya 705 data. Dataset ini mencakup beberapa kolom, yaitu*, Username, Score, At, Content* yang mana data ini akan digunakan untuk mengklasifikasikan sentimen ulasan sebagai sentimen positif atau negatif menggunakan algoritma *Naïve Bayes Classifier.*

Tabel 4. 1 Hasil Selection

No	Tgl	Nama	Rant	Review
1.	25/11/2024 09:25	KP Inenga Stephi astuti	5	terperinci dan cepat
2.	25/11/2024 09:16	Aries Munandar	5	tlg ditambahkan opsi pembayaran VA via BCA.jd pengguna BCA tdk hrs keluar biaya ekstra .sukses selalu!!
	...	...	...	...
704	25/11/2024 09:12	Joko Priyanto	5	Pelayanannya bagus
705	25/11/2024 09:12	Risya El Haris	5	alhamdulillah .. adanya ada kami ,kami sangat terbantu .. dan proses sangat cepat

4.1.2 Preprocessing Data

Dalam bagian ini, kami akan melakukan beberapa langkah utama dalam pra pemrosesan data sebelum melatih model algoritma *Naive Bayes*. Proses ini mencakup penghapusan data yang tidak relevan, pembersihan teks, dan transformasi data menjadi format yang dapat diproses oleh model pembelajaran mesin.

a.Cleaning teks

Langkah-langkah ini bertujuan untuk memastikan bahwa teks dalam dataset bebas dari elemen yang tidak relevan, sehingga siap untuk analisis lebih lanjut seperti tokenisasi, *stemming*, atau pembelajaran mesin.

Tabel 4. 2 Hasil Cleaning

Asal	Hasil
tlg ditambahkan opsi pembayaran VA via BCA.jd pengguna BCA tdk hrs keluar biaya ekstra. sukses selalu!!	tlg ditambahkan opsi pembayaran va via bca.jd pengguna bca tdk hrs keluar biaya ekstra. sukses selalu!!

b.Normalisasi

Normalisasi adalah proses yang digunakan untuk mengubah kata-kata tidak baku atau kata-kata yang tidak sesuai dengan kaidah bahasa menjadi bentuk yang baku[4]. Prosess ini membantu meningkatkan kualitas data teks dengan memastikan konsistensi kata-kata, sehingga lebih siap untuk langkah analisis lebih lanjut seperti tokenisasi atau pemodelan pembelajaran mesin.

Tabel 4. 3 Hasil Normalisasi

Asal	Hasil
tlg ditambahkan opsi pembayaran va via bca.jd pengguna bca tdk hrs keluar biaya ekstra. sukses selalu!!	tolong ditambahkan opsi pembayaran va via bca.jd pengguna bca tidak harus keluar biaya ekstra.sukses selalu!!

c.Tokenizing

Proses tokenisasi ini membagi teks menjadi kata-kata terpisah, yang kemudian dapat digunakan untuk langkah-langkah analisis teks lebih lanjut seperti pemrosesan dan analisis frekuensi kata, pembuatan fitur untuk model pembelajaran mesin, dan lainnya.

Tabel 4. 4 Hasil Tokenizing

Asal	Hasil
tolong ditambahkan opsi pembayaran va via bca.jd pengguna bca tidak harus keluar biaya ekstra.sukses selalu!!	['tolong', 'ditambahkan', 'opsi', 'pembayaran', 'va', 'via', 'bca.jd', 'pengguna', 'bca', 'tidak', 'harus', 'keluar', 'biaya', 'ekstra', sukses', 'selalu', '!', '!']

d.Stopword
langkah-langkah ini penting untuk merampingkan teks, sehingga hanya kata-kata yang relevan tetap ada, yang kemudian dapat digunakan untuk analisis lebih lanjut seperti ekstraksi fitur atau pemodelan pembelajaran mesin.

Tabel 4. 5 Hasil Stopword

Asal	Hasil
tolong ditambahkan opsi pembayaran va via bca.jd pengguna bca tidak harus keluar biaya ekstra.sukses selalu!!	['tolong', 'opsi', 'pembayaran', 'va', 'via', 'bca.jd', 'pengguna', 'bca', 'tdk', 'biaya', 'ekstra', sukses', '!', '!']

e.Steming
Proses stemming ini mengurangi variasi kata-kata yang memiliki arti sama, sehingga teks lebih konsisten. Langkah ini sangat penting untuk analisis teks lebih lanjut, seperti pemodelan pembelajaran mesin atau pengelompokan dokumen.

Tabel 4. 6 Hasil Stemming

Asal	Hasil
tolong ditambahkan opsi pembayaran va via bca.jd pengguna bca tidak harus keluar biaya ekstra.sukses selalu!!	['tolong', 'opsi', 'bayar', 'va', 'via', 'bca jd', 'guna', 'bca', 'tdk', 'biaya', 'ekstra', sukses', '', '']

4.1.3 Data Tranformation

Transformasi Data adalah proses mengubah data dari format atau struktur aslinya menjadi format yang lebih sesuai untuk analisis atau pemrosesan lebih lanjut. Proses ini mencakup langkah-langkah seperti normalisasi, pengkodean, penskalaan, atau penggabungan data, dengan tujuan meningkatkan kualitas dan kompatibilitas data untuk mendukung hasil analisis yang akurat.

a.Lebeling
Langkah ini memungkinkan klasifikasi data berdasarkan sentimen (positif atau negatif) untuk analisis lebih lanjut. Pengelompokan berdasarkan sentimen berguna untuk memahami distribusi persepsi pengguna terhadap suatu produk, layanan, atau konten.
b.TF-IDF
Proses ini menghasilkan representasi numerik teks yang dapat digunakan sebagai input untuk algoritma pembelajaran mesin. Representasi *TF-IDF* membantu mempertahankan relevansi kata dalam dokumen dengan mempertimbangkan frekuensi kemunculan kata dalam dataset secara keseluruhan.

4.1.4 Data Transformation

Transformasi Data adalah proses mengubah data dari format atau struktur aslinya menjadi format yang lebih sesuai untuk analisis atau pemrosesan lebih lanjut. Proses ini mencakup langkah-langkah seperti normalisasi, pengkodean, penskalaan, atau penggabungan data, dengan tujuan meningkatkan kualitas dan kompatibilitas data untuk mendukung hasil analisis yang akurat.

a. Labeling

Langkah ini memungkinkan klasifikasi data berdasarkan sentimen (positif atau negatif) untuk analisis lebih lanjut. Pengelompokan berdasarkan sentimen berguna untuk memahami distribusi persepsi pengguna terhadap suatu produk, layanan, atau konten.

label	count
positif	457
negatif	248

Gambar 2.Hasil Lebeling

b.TF-IDF

Proses ini menghasilkan representasi numerik teks yang dapat digunakan sebagai input untuk algoritma pembelajaran mesin. Representasi *TF-IDF* membantu mempertahankan relevansi kata dalam dokumen dengan mempertimbangkan frekuensi kemunculan kata dalam dataset secara keseluruhan.

	acc	ada	adakami	aja	akan	aplikasi	bagus	baik	banget	banyak	...	terimakasih	terpercaya	terus	tidak	tolak	tolong	uang	udah
0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0
1	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.210589	0.0	0.276912	0.0	0.0
2	0.0	0.000000	0.000000	0.0	0.0	0.270509	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0
3	0.0	0.000000	0.000000	0.0	0.0	0.000000	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0
4	0.0	0.264978	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
700	0.0	0.000000	0.505477	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0
701	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0
702	0.0	0.000000	0.000000	0.0	0.0	0.161086	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.187250	0.0	0.000000	0.0	0.0
703	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0
704	0.0	0.000000	0.000000	0.0	0.0	0.545021	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0

705 rows × 83 columns

Gambar 3. Hasil TF-IDF

4.1.5 Data Mining

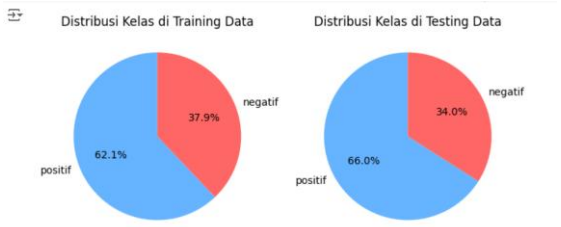
Data Mining adalah proses menganalisis dan mengekstraksi pola, informasi, atau pengetahuan yang berguna dari kumpulan data yang besar dan kompleks. Proses ini menggunakan teknik statistik, pembelajaran mesin, dan algoritma untuk mengidentifikasi tren, hubungan, atau anomali yang tidak terlihat secara langsung, dengan tujuan mendukung pengambilan keputusan yang lebih baik.

a.Spliting Data

Proses di atas melakukan pembagian dataset menjadi data latih dan data uji, kemudian mengubah teks menjadi representasi numerik menggunakan teknik *TF-IDF,* dengan tujuan mempersiapkan data untuk digunakan dalam model pembelajaran mesin. Model nantinya akan dilatih menggunakan data latih *(X_train_tfidf)* dan diuji menggunakan data uji *(X_test_tfidf).*

b.Training dan Testing

Kode ini membagi data menjadi *training* dan *testing* set, menghitung distribusi kelas pada masing-masing set, dan menampilkan dua *pie chart* untuk memvisualisasikan distribusi kelas tersebut.



Gambar 4. Hasil Testing dan Training

d. NAIVE BAYES dan SMOTE

ini melakukan transformasi teks menggunakan *TF-IDF*, menyeimbangkan data latih dengan *SMOTE*, melatih model *Naive Bayes* pada data latih yang diseimbangkan, mengevaluasi performa model menggunakan akurasi, confusion matrix, dan laporan klasifikasi, serta menambahkan hasil prediksi model ke dalam *DataFrame*.

```
Akurasi Train: 0.97
Akurasi Test: 0.83

Confusion Matrix:
[[142  26]
 [ 58 268]]

Classification Report:
              precision    recall  f1-score   support

     negatif       0.71      0.85      0.77       168
     positif       0.91      0.82      0.86       326

    accuracy                           0.83       494
   macro avg       0.81      0.83      0.82       494
weighted avg       0.84      0.83      0.83       494
```

Gambar 5. Hasil NAÏVE BAYES

4.1.6 Evaluasi

a.Cross Validation

teknik evaluasi model yang digunakan untuk mengukur performa dan generalisasi model terhadap data baru dengan membagi dataset menjadi beberapa subset atau lipatan (folds). Dalam proses ini, model dilatih pada sebagian data (data latih) dan diuji pada bagian lainnya (data uji) secara bergantian hingga semua subset digunakan. Metode ini membantu mengurangi bias dan memastikan bahwa model tidak hanya bekerja baik pada data latih tetapi juga pada data yang belum pernah dilihat, sehingga menghasilkan evaluasi yang lebih akurat dan andal.

```
Cross-Validation Scores: [0.83018868 0.96226415 0.90384615 0.86538462 0.92307692]
Mean Accuracy: 0.8969521044992744
Standard Deviation: 0.045715552700324645
```

Gambar 6. Hasil Cross Validation

4.2 HASIL

4.2.1.Accuracy

Pada model analisis sentimen menggunakan algoritma *Naïve Bayes* , metrik akurasi digunakan untuk mengukur seberapa sering model membuat prediksi yang benar secara keseluruhan. Metrik ini penting untuk menilai kemampuan model dalam membedakan sentimen positif dan negatif pada ulasan pengguna aplikasi Adakami. Akurasi yang tinggi menunjukkan bahwa model mampu mengklasifikasikan ulasan dengan tepat pada tingkat keseluruhan yang berarti model dapat secara efektif memisahkan ulasan positif dari ulasan negatif. Berikut ini adalah cara menghitung akurasi model menggunakan matriks kebingungan *(confusion matrix)*.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{285 + 144}{285 + 144 + 25 + 48} = \frac{429}{502} = 0.85\ (85\%)$$

Gambar 7. Rumus Accuracy

Penjelasan rumus *(Accuracy)* :

TP (True Positives) : 285 positif klasifikasi benar.

TN (True Negative) : 144 negatif klasifikasi benar.

FP(FalsePositives) : 25 negatif klasifikasi sebagai positif.

FN(FalseNegatives) : 48 positif klasifikasi sebagai negatif.

Akurasi sebesar 85% menunjukkan bahwa model berhasil memprediksi sentimen ulasan dengan benar pada 85% dari total data uji. Ini berarti bahwa model memiliki performa yang cukup baik dalam membedakan ulasan positif dan negatif,

4.2.2. Precision

Metrik *precision* mengukur sejauh mana prediksi positif yang dibuat oleh model benar-benar positif. *Precision* dihitung menggunakan rumus:

$$Precision = \frac{TP}{TP + FP} = \frac{285}{285 + 25} = \frac{267}{289} = 0.92\ (92\%)$$

Gambar 8. Rumus Precision

Penjelasan rumus *(Precision)* :

TP (True Positives) : 285 jumlah ulasan positif yang diklasifikasikan benar oleh model.

FP (False Positives) : 25 jumlah ulasan negatif yang salah diklasifikasikan ulasan positif.

Precision sebesar 92% menunjukkan bahwa dari semua prediksi positif yang dihasilkan oleh model, 92% di antaranya benar-benar positif. Ini berarti model cukup andal dalam mengidentifikasi ulasan positif dengan sedikit kesalahan prediksi.

4.2.3. Recall

Matrik Recall mengukur seberapa baik model dalam menemukan semua ulasan positif yang ada. *Recall* dihitung menggunakan rumus:

$$Recall = \frac{TP}{TP + FP} = \frac{285}{285 + 48} = \frac{285}{333} = 0.86\ (86\%)$$

Gambar 9. Rumus Recall

Penjelasan rumus *(Recall)* :

TP (True Positives) : 285 jumlah ulasan positif yang diklasifikasikan dengan benar.

FN (False Negatives) : 48 jumlah ulasan positif salah diklasifikasikan sebagai ulasan negatif.

Recall sebesar 86% menunjukkan bahwa model mampu mengidentifikasi 86% dari total ulasan positif yang ada. Namun, sebanyak 14% ulasan positif terlewat (salah diklasifikasikan sebagai negatif).

4.2.4.F1-Score

Matrik *F1-Score* digunakan untuk menggabungkan precision dan recall dalam satu nilai dengan rata-rata harmonis. *F1-Score* dihitung menggunakan rumus :

$$F1\text{-}Score\ 2\ x = \frac{Precision\ x\ Recall}{Precision + Recall} = 2\ x\ \frac{0.92\ x\ 0.86}{0.92 + 0.86} = 2\ x\ \frac{0.7912}{1.78} = 0.89\ (89\%)$$

Gambar 10. Rumus F1-Score

Penjelasan rumus *(F1 - Score)* :

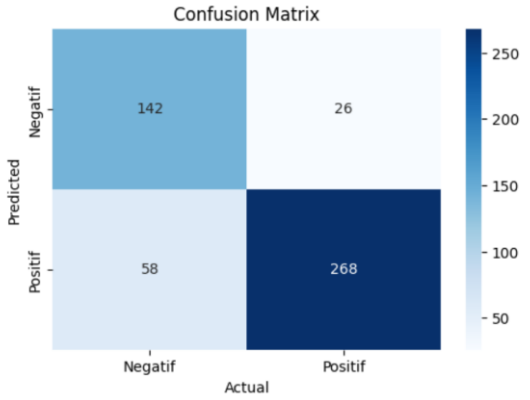
Precision menunjukkan 0.92 klasifikasi positif yang benar-benar positif.

Recall menunjukkan 0.86 ulasan positif yang teridentifikasi dengan benar.

F1-Score menghasilkan nilai 89% menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Berdasarkan hasil perhitungan *F1-Score* yang menunjukkan model ini akurat dalam mengklasifikasi ulasan positif dengan tepat.

4.2.5. Confusion Matrix

Confusion matrix adalah tabel yang merangkum hasil klasifikasi yang benar dan salah yang dihasilkan oleh model serta menunjukkan distribusi klasifikasi untuk setiap kelas. Tabel ini sangat penting untuk memahami bagaimana model dalam analisis sentimen. Oleh karena itu, analisis matriks kebingungan memungkinkan kita untuk mengevaluasi dan memperbaiki kinerja model dalam membedakan antara sentimen positif dan negatif. Berikut adalah tahapan kode untuk menampilkan visualisasi *confusion matrix.*

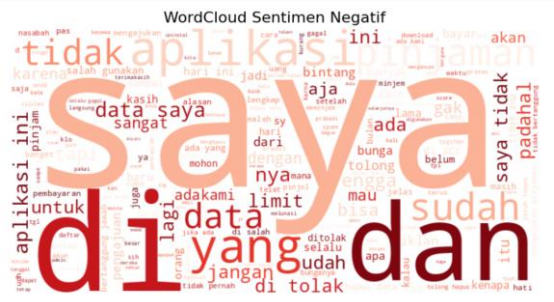


Gambar 11. Hasil Confusion Matrix

4.2.6. Word Cloud

Word Cloud merupakan representasi visual dari kumpulan teks dimana frekuensi atau kepentingan suatu kata ditampilkan melalui ukuran dan posisi kata tersebut. Kata-kata yang muncul lebih sering dalam teks.Berikut kode *WordCloud* Positif dan Negatif :

Gambar 12. Hasil Word Cloud Positif



Gambar 13. Hasil Word Cloud Negatif

5. KESIMPULAN

a. sentimen pengguna terhadap aplikasi AdaKami di Google Play Store total dataset berjumlah 1000

data,Di sleksi menjadi 705 data ditemukan jumlah ulasan sentimen positif sebanyak 457 data sedangkan jumlah ulasan sentimen positif sebanyak 258 data. Hasil analisis sentimen pengguna terhadap aplikasi AdaKami di Google Play Store didominasi oleh sentimen positif.

b. Hasil evaluasi dari algoritma Multinomial Naïve Bayes yang telah dilakukan dalam mengklasifikasikan sentimen ulasan pengguna aplikasi AdaKami menggunakan NAIVE BAYES mendapatkan hasil evaluasi dengan nilai akurasi 85%, presisi 92%, recall 86%, dan f1-score 89%. Model bekerja cukup baik secara keseluruhan. Model sangat akurat dalam memprediksi kelas positif, tetapi mungkin mengorbankan recall. Model mendeteksi sebagian besar data positif, tetapi ada beberapa data positif yang terlewat. Kombinasi precision dan recall menunjukkan performa yang baik, dengan keseimbangan yang mendekati ideal.
c. Hasil cross-validation memberikan skor akurasi untuk setiap lipatan (fold) dalam proses k-fold cross-validation. Dalam hal ini, model Anda dievaluasi pada 5 fold, dengan hasil 0.8302 , 0.9623, 0.9038, 0.8654, 0.9231.Mean Accuracy: 0.8969 (89.69%). Standard Deviation: 0.0457 (4.57%). Model ini memiliki kinerja rata-rata yang baik (89.69%) dengan variasi yang kecil di antara fold, menandakan model stabil dan dapat diandalkan.
d. Bisa di simpulkan dari Word Cloud bahwa aplikasi ada kami sangat membatu dan pencairan nya cepat tetapi ada juga keluhan terkait di tolak saat pengajuan pinjaman, bunga terlalu tinggi, dan keluhan lainnya. Maka saran yang dapat diterapkan bagi penelitian selanjutnya yaitu, penggunaan algoritma lainnya untuk bisa menemukan nilai evaluasi yang lebih baik. Coba gunakan regularisasi untuk mengurangi overfitting. Pastikan dataset pelatihan dan pengujian seimbang.Eksperimen dengan teknik data augmentation atau hyperparameter tuning.

UCAPAN TERIMA KASIH

Terima kasih dan bersyukur kepada semua pihak yang telah memberikan dukungan dan berperan aktif dalam penelitian ini.

1. Bapak Assoc. Prof. Dr. Dadang Sudrajat, S.Si., M.Kom, selaku Ketua STMIK IKMI Cirebon.
2. Bapak Dian Ade Kurnia, M.Kom, selaku Wakil Ketua I Bidang Akademik, Kerjasama, Riset dan Inovasi.
3. Ibu Dra. Nining R, M.Si., selaku Wakil Ketua II Bidang Keuangan.
4. Ibu Fatihanursari Dikananda, S.Tr.I.Kom., M.Kom selaku Wakil Ketua III Bidang Kemahasiswaan dan Alumni.
5. Bapak H. Eka Jayawangsa, BBA., selaku Wakil Ketua IV Bidang Sarana dan Prasarana.
6. Ibu Gifthera Dwilestari, S.I.Kom., M.Kom, Sebagai Ketua Program Studi Teknik Informatika.
7. Bapak Bambang Irawan, M.T., sebagai Dosen Pembimbing Utama.
8. Bapak Willy Prihartono, M.kom., sebagai dosen Pembimbing Kedua.

DAFTAR PUSTAKA

[1] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis Sentimen Ulasan Pengguna Aplikasi Samsat Digiital Nasional (Signal) Dengan Menggunakan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 963–969, 2024, doi: 10.23960/jitet.v12i2.4073.

[2] "Eskiyaturrofikoh" and R. R. 'Suryono, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm)," *JIPI(Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1408–1419, 2024, [Online]. Available: https://www.jurnal.stkippgritulungagung.ac.id

/index.php/jipi/article/view/5392

[3] M. R. Alta Zahir, B. Ramdani, and A. Dwi Saputra, "Analisis Sentimen Terhadap Ulasan Aplikasi Pinjaman Online (PINJOL) di Google Play Store Menggunakan Naive Baiyes Classifier," *J. Ris. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 61–64, 2024, doi: 10.58776/jriti.v1i2.39.

[4] N. Z. Rania and R. D. Syah, "Analisis Sentimen Terhadap Aplikasi Gojek Pada Play Store Menggunakan Metode Random Forest Classifier," *J. Ilm. Inform. Komput.*, vol. 29, no. 2, pp. 144–153, 2024, doi: 10.35760/ik.2024.v29i2.11877.

[5] A. Nuraini, A. Faqih, G. Dwilestari, N. Dienwati Nuris, and R. Narasati, "Analisis Sentimen Terhadap Review Aplikasi Brimo Di Google Play Store Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3661–3666, 2024, doi: 10.36040/jati.v7i6.8228.

[6] T. Hartati, R. T. Sohadi, E. Tohidi, and E. Wahyudin, "Penerapan Algoritma Naive Bayes pada Analisis Sentimen Ulasan Aplikasi Whoosh – Kereta Cepat Di Google Play Store," vol. 6, no. 1, pp. 244–249, 2024.

[7] D. Surya Sayogo, B. Irawan, and A. Bahtiar, "Analisis Sentimen Ulasan Instagram Di Google Play Store Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3314–3319, 2024, doi: 10.36040/jati.v7i6.8178.

[8] M. Tirta Nugraha, N. Nina Sulistiyowati, and U. Ultach Enri, "Analisis Sentimen Ulasan Aplikasi Satu Sehat Pada Google Play Store Menggunakan Naïve Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 5, pp. 3593–3601, 2024, doi: 10.36040/jati.v7i5.7753.

[9] A. Nurian, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: 10.23960/jitet.v11i3s1.3348.

[10] N. Fibriyanti Arminda, N. Sulistiyowati, and T. Nur Padilah, "Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, pp. 1817–1822, 2023, doi: 10.36040/jati.v7i3.7012.

[11] I. K. J. Arta, G. Indrawan, and G. Rasben Dantes, "Data Mining Rekomemdasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode Technique For Other Reference By Similarity to Ideal Solution," *J. Ilmu Komput. Indones.*, vol. 4, no. 1, pp. 11–21, 2019.

[12] T. Asy Aria, M. Julkarnain, and F. Hamdani, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Algoritma K-Means Clustering Untuk Data Obat," *Media Online*, vol. 4, no. 1, pp. 649–657, 2023, doi: 10.30865/klik.v4i1.1117.

[13] I. Nursatika Kusuma and I. Ali, "Analisis Sentimen Pada Pengguna Aplikasi Dana Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1470–1476, 2024, doi: 10.36040/jati.v8i2.9041.

[14] D. D. A. Yani, H. S. Pratiwi, and H. Muhardi, "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.

[15] N. Wijaya and E. S. Panjaitan, "Analisis Sentimen Ulasan Aplikasi Instagram di Google Play Store : Pendekatan Multinomial Naive Bayes dan Berbasis Leksikon," vol. 6, no. 2, pp. 921–929, 2024, doi: 10.47065/bits.v6i2.5615.