PEMANFAATAN *MULTINOMIAL NAIVE BAYES* UNTUK ANALISIS SENTIMEN ULASAN APLIKASI *MOBILE* JKN

Isra Elwanda Putra, Asriyanik, Fathia Frazna Azzahra
Teknik Informatika, Universitas Muhammadiyah Sukabumi
Jl. R. Syamsudin, S.H. No. 50, Cikole, Kec. Cikole, Kota Sukabumi, Jawa Barat 43113
israelwandateater@gmail.com

ABSTRAK

Aplikasi kesehatan digital seperti *Mobile* JKN menjadi solusi penting dalam layanan medis, terutama selama pandemi COVID-19. Namun, ulasan pengguna di Google Play Store mengungkapkan masalah, seperti *bug*, respons lambat, dan fitur yang belum optimal, yang dapat mengurangi kepercayaan pengguna. Untuk itu, penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi *Mobile* JKN guna memahami persepsi mereka dan mengidentifikasi aspek-aspek yang perlu ditingkatkan. Penelitian ini mengadopsi pendekatan KDD (*Knowledge Discovery in Databases*), yang meliputi tahapan *data selection*, *pre-processing*, transformasi data menggunakan TF-IDF dan *Chi-square*, serta analisis sentimen dengan algoritma *Multinomial Naïve Bayes*. Algoritma ini dipilih karena kemampuannya dalam menangani klasifikasi teks pada dataset besar dan kompleks. Dataset terdiri dari ulasan pengguna *Mobile* JKN, yang diklasifikasikan ke dalam sentimen positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa model mencapai akurasi 85,5%, dengan performa yang sangat baik dalam mengidentifikasi ulasan negatif (*recall* tinggi) serta prediksi negatif dan positif yang akurat (*precision* optimal). Namun, model kurang efektif dalam mengenali ulasan netral dan mendeteksi ulasan positif, kemungkinan akibat representasi data yang belum optimal. Oleh karena itu, diperlukan perbaikan dalam *pre-processing* dan pemilihan fitur untuk meningkatkan keseimbangan klasifikasi. Temuan ini menegaskan bahwa pemilihan fitur yang tepat dan strategi *pre-processing* yang optimal dalam meningkatkan performa model analisis sentimen.

Kata kunci : *Analisis Sentimen, Mobile JKN, KDD, TF-IDF, Chi-Square, Multinomial Naïve Bayes*

1. PENDAHULUAN

Aplikasi kesehatan digital, seperti *Mobile* JKN, memegang peranan krusial, terutama sejak pandemi COVID-19, dengan 57% masyarakat Indonesia beralih menggunakan aplikasi Kesehatan [1].

Mobile JKN merupakan aplikasi dari BPJS Kesehatan yang memfasilitasi akses layanan kesehatan, khususnya bagi pasien rumah sakit, melalui fitur pendaftaran online, pengecekan ketersediaan kamar, dan lainnya [2].

Meskipun popularitasnya meningkat, *Mobile* JKN menghadapi tantangan terkait fungsionalitas, kinerja, dan kepuasan pengguna, yang tercermin dari ulasan di platform seperti Google Play Store. Permasalahan ini dapat menghambat akses layanan kesehatan bagi pasien yang sangat bergantung pada aplikasi ini. Jika tidak segera diatasi, tingkat kepercayaan pengguna dapat menurun, yang berpotensi memperlambat adopsi *Mobile* JKN di masa mendatang. Oleh karena itu, evaluasi ulasan pengguna menjadi penting untuk memahami persepsi dan pengalaman mereka. Berbagai upaya telah dilakukan untuk meningkatkan kualitas aplikasi, termasuk pengumpulan dan analisis ulasan serta penerapan analisis sentimen berbasis *machine learning*.

Penelitian sebelumnya menunjukkan bahwa Naïve Bayes(NB) dengan proporsi data *training* 70% dan *testing* 30% menghasilkan akurasi tertinggi (94,75%). Analisis tersebut juga menemukan bahwa lebih dari 70% dari 10.424 komentar pengguna Mobile JKN bersentimen positif [3].

Penelitian ini memberikan kontribusi yang berbeda dibandingkan penelitian sebelumnya dalam tiga aspek utama:

a. penggunaan dataset yang lebih besar dan beragam, yang mencakup variasi ulasan pengguna yang lebih representatif
b. eksplorasi berbagai proporsi pembagian data latih dan uji untuk menguji robustitas model
c. implementasi kombinasi algoritma Multinomial Naive Bayes (MNB) dengan metode lain dengan tujuan meningkatkan akurasi dan konsistensi hasil analisis sentimen.

Penelitian ini, yang berjudul "Pemanfaatan *Multinomial Naive Bayes* dan TF-IDF untuk Analisis Sentimen Ulasan Aplikasi *Mobile* JKN", bertujuan mengeksplorasi optimalisasi performa model untuk menghasilkan analisis sentimen yang lebih akurat dan tepat sasaran.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen, atau *opinion mining*, berfokus pada identifikasi dan klasifikasi opini atau perasaan dalam teks ke dalam polaritas positif, negatif, atau netral. Berbeda dengan fakta yang objektif, opini bersifat subjektif dan mencerminkan perasaan atau persetujuan. Terdapat dua pendekatan utama dalam analisis sentimen: berbasis *lexicon* (menggunakan kamus kata dengan nilai polaritas) dan *machine learning* (menggunakan algoritma untuk mengenali pola). Dibandingkan metode *lexicon*, pendekatan

machine learning, khususnya *supervised learning* dan *unsupervised learning*, umumnya memberikan hasil yang lebih akurat dan efisien [4].

2.2. *Mobile* JKN

Mobile JKN merupakan aplikasi dari BPJS Kesehatan yang memfasilitasi akses berbagai layanan bagi peserta melalui *smartphone*. Fitur-fiturnya meliputi pendaftaran, perubahan dan pengecekan data kepesertaan, akses fasilitas kesehatan (FKTP dan FKTL), serta penyampaian kritik dan saran [5].

2.3. *Website*

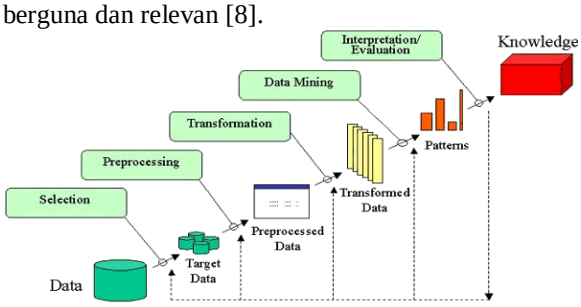
Website adalah sekumpulan halaman yang menyajikan beragam informasi, termasuk teks, gambar (baik statis maupun dinamis), animasi, serta suara, atau gabungan dari elemen-elemen tersebut dalam bentuk yang interaktif. Halaman-halaman ini saling terhubung dalam suatu struktur yang berkesinambungan melalui jaringan antar halaman [6].

2.4. *Machine Learning*

Machine Learning (ML) adalah cabang ilmu komputer yang mengembangkan algoritma untuk belajar dari data dan meningkatkan kinerjanya secara otomatis. Teknologi ini memungkinkan komputer mengenali pola, membuat prediksi, dan mengambil keputusan tanpa instruksi eksplisit. ML membangun model dari data pelatihan agar komputer dapat beradaptasi dan berkembang seiring waktu [7].

2.5. *Knowledge Discovery in Databases*

Knowledge Discovery in Databases (KDD) adalah proses sistematis untuk mengekstraksi informasi dari basis data. Seiring dengan kemajuan teknologi, kebutuhan akan penyimpanan data dalam jumlah besar semakin meningkat. Oleh karena itu, basis data harus memiliki kapasitas yang memadai untuk dianalisis guna menghasilkan informasi yang berguna dan relevan [8].



Gambar 1. KDD *Process*

[9] menyatakan bahwa proses *Knowledge Discovery in Databases* (KDD) terdiri dari beberapa tahapan utama, yaitu:

a. *Data Selection*
Sebelum proses KDD dimulai, data operasional harus diseleksi terlebih dahulu. Data yang telah dipilih untuk proses *data mining* kemudian disimpan dalam file terpisah dari basis data operasional.

b. *Pre-processing/Cleaning*
Sebelum *data mining*, data perlu dibersihkan, termasuk penghapusan data duplikat, pemeriksaan konsistensi, dan perbaikan kesalahan. Data juga dapat diperkaya dengan informasi eksternal. Khusus untuk data teks, *text pre-processing* penting untuk meningkatkan akurasi model *machine learning*. Tahapannya meliputi *case folding* (menjadi huruf kecil), *stopword filtering* (menghapus kata tidak penting), *tokenizing* (memisahkan kalimat menjadi kata), dan *stemming* (menghilangkan imbuhan). Selain itu, konversi singkatan atau kata *slang* menjadi kata standar juga dapat dilakukan [10].

c. *Transformation*
Dalam proses KDD, TF-IDF berada pada tahap transformasi, di mana teks dikonversi menjadi vektor numerik. Metode ini menghitung bobot kata berdasarkan frekuensi dalam dokumen (TF) dan kelangkaannya di dokumen lain (IDF), sehingga menyoroti istilah yang paling relevan. Representasi ini kemudian digunakan sebagai input untuk algoritma klasifikasi guna meningkatkan akurasi analisis [10].
Dalam KDD, metode *Chi-Square* digunakan pada tahap transformasi sebagai teknik *supervised feature selection* untuk menghilangkan fitur yang tidak relevan tanpa menurunkan akurasi klasifikasi dokumen. Metode ini mengukur hubungan antara term dan kategori sentimen menggunakan teori statistika, lalu memilih fitur dengan ketergantungan tinggi untuk meningkatkan akurasi model. Dengan demikian, *Chi-Square* membantu membuat klasifikasi lebih efisien dan fokus pada fitur yang paling relevan [11].

d. *Data mining*
Data mining adalah proses semi-otomatis yang menggunakan metode statistik, matematika, AI, dan *machine learning* untuk menemukan informasi tersembunyi dalam *database* besar. [12].
Multinomial Naïve Bayes (MNB) adalah metode klasifikasi khusus untuk data teks. Keunggulannya terletak pada pendekatan independen, di mana setiap dokumen diklasifikasikan secara terpisah berdasarkan isinya tanpa dipengaruhi oleh dokumen lain [13]. [14] menyatakan bahwa rumus MNB dinyatakan sebagai:

$$P(c|d) = \frac{N_c}{N} \times \prod_{i=1}^{n} P(t_i|c) \quad (1)$$

Keterangan:
$P(c|d)$: Probabilitas suatu dokumen (d) termasuk dalam kelas (c).
N_c : Jumlah kelas (c) pada seluruh dokumen.
N : Jumlah seluruh dokumen.
$\prod_{i=1}^{n} P(t_i|c)$: Probabilitas kata ke-i sampai n dengan diketahui kelas (c).

e. *Interpretation* & *Evaluation*
Pada tahap ini, evaluasi model dilakukan menggunakan *confusion matrix*, yang menilai kinerja model dengan membandingkan nilai sebenarnya (*true value*) dan prediksi. Matriks ini mengidentifikasi data yang diklasifikasikan dengan benar atau salah serta menghasilkan metrik seperti akurasi, presisi, *recall*, dan *F1-measure*. *Confusion matrix* terdiri dari empat komponen utama: *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN) [15].
Hasil dari *confusion matrix* digunakan untuk menghitung nilai akurasi, presisi, dan *recall* [14].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

2.6. *Web Scraping*

Web scraping adalah teknik otomatis untuk mengekstrak data dari halaman *web* berformat HTML atau XHTML. Metode ini digunakan untuk mengumpulkan informasi tanpa perlu menyalin data secara manual, sehingga mempermudah pencarian data spesifik dan terarah [16].

3. METODE PENELITIAN

3.1. Identifikasi Masalah

Penelitian ini membahas tantangan aplikasi *Mobile* JKN dalam memenuhi ekspektasi pengguna terkait fungsionalitas, kinerja, dan kepuasan. Meskipun mempermudah akses layanan kesehatan, pengguna mengeluhkan respons lambat, *bug*, dan fitur yang kurang optimal. Melalui analisis sentimen dengan *Multinomial Naïve Bayes* dan TF-IDF, penelitian ini bertujuan mengkaji ulasan pengguna guna mengembangkan model yang lebih akurat.

3.2. Studi Literatur

Setelah identifikasi masalah, studi literatur dilakukan untuk memahami analisis sentimen, termasuk *Multinomial Naïve Bayes* dan TF-IDF. Penelitian sebelumnya ditinjau untuk mengatasi tantangan seperti pemrosesan teks, ketidakseimbangan kelas, dan performa model. Selain itu, keunggulan *Multinomial Naïve Bayes* dibandingkan algoritma lain diperkuat sebagai dasar pemilihan metode utama penelitian.

3.3. *Data Selection*

Data ulasan aplikasi *Mobile* JKN dikumpulkan dari Play Store menggunakan pustaka google-play-scraper, mencakup nama pengguna, tanggal, skor, dan isi ulasan. Ulasan dalam Bahasa Indonesia yang relevan dipilih, sementara ulasan yang terlalu singkat atau tidak relevan disaring untuk memastikan kualitas dataset.

3.4. *Pre-processing*

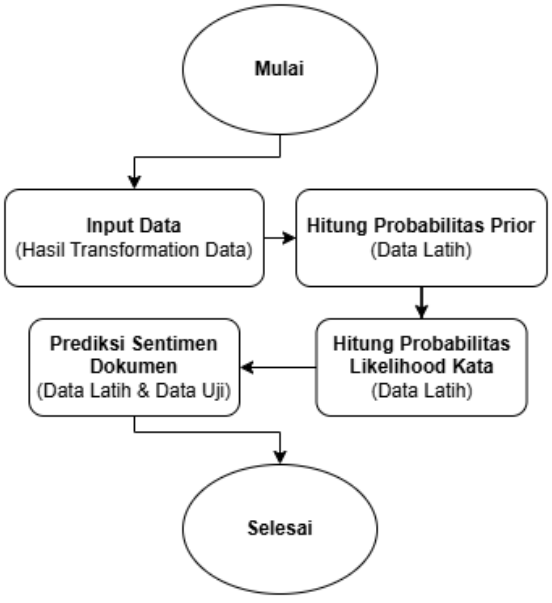
Data yang terkumpul kemudian melalui proses *Filtering & Cleansing* untuk memastikan kualitasnya. Langkah ini meliputi penghapusan ulasan duplikat, penanganan data kosong, serta penghilangan kata-kata, karakter, atau simbol yang tidak relevan.

Setelah itu, dilakukan *text pre-processing* untuk mempersiapkan teks. Proses dimulai dengan *case folding* untuk mengubah teks menjadi huruf kecil. Kemudian, teks dipecah menjadi kata-kata melalui tokenisasi, diikuti dengan normalisasi untuk mengganti kata slang atau tidak baku menjadi kata baku. Kata-kata yang tidak penting, seperti *stopwords*, dihapus menggunakan daftar *stopwords*. Proses terakhir adalah *stemming* untuk mengembalikan kata ke bentuk dasar.

3.5. *Transformation*

Pada tahap ini, teks diubah menjadi representasi numerik dengan menggunakan teknik TF-IDF, yang dihitung berdasarkan *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Selanjutnya, seleksi fitur dilakukan menggunakan metode *Chi-Square* untuk menilai hubungan antara fitur (kata) dan label (sentimen). Fitur dengan skor *Chi-Square* tinggi dipilih untuk membentuk dataset yang lebih kecil namun tetap informatif.

3.6. *Data Mining*



Gambar 2. Alur *Data Mining*

Setelah transformasi data, tahap *data mining* dimulai dengan penerapan algoritma *Multinomial Naive Bayes* (MNB) untuk analisis. Model MNB dilatih menggunakan dataset pelatihan untuk mempelajari probabilitas sentimen berdasarkan distribusi kata. Kemudian, model ini digunakan untuk memprediksi sentimen pada data pengujian. MNB dipilih karena keefisienannya dalam klasifikasi teks.

3.7. *Interpretation & Evaluation*

Setelah model diterapkan, hasilnya dievaluasi menggunakan metrik performa seperti akurasi, *precision*, *recall*, dan *f1-score* untuk menilai efektivitas model. Selain itu, evaluasi juga dilakukan dengan menggunakan *confusion matrix* untuk memberikan gambaran lebih detail tentang kinerja model dalam mengklasifikasikan data.

3.8. Visualisasi

Hasil analisis sentimen ditampilkan melalui *website* interaktif yang memungkinkan pengguna memasukkan ulasan dan mendapatkan hasil analisis instan. Informasi yang ditampilkan mencakup tingkat kepercayaan, teks asli, teks yang telah diproses, dan klasifikasi sentimen.

3.9. Kesimpulan

Pada tahap ini, penelitian disimpulkan dengan merangkum temuan utama, menjawab pertanyaan penelitian, dan memberikan rekomendasi untuk penelitian selanjutnya atau penerapan praktis.

4. HASIL DAN PEMBAHASAN

4.1. *Data Selection*

Tabel 1. Contoh data *scraping* setelah diseleksi

Score	*Content*
1	Saya mau daftar .. muncul No hp sudah digunakan.. Saya baru pertama kali pake aplikasi nya
4	Sangat membantu sekali.salam sehat dan sukses selalu kedepannya.ðŸ™□ðŸ™□ðŸ‘□ðŸ‘□ðŸ¤²ðŸ¤²
1	Ini aplikasi kenapa buruk sekali sistemnya. Verifikasi nomer telpon butuh waktu ratusan detik tapi tetap tidak ada. Bukannya memudahkan malah menyusahkan.
5	Buat yang gk masuk kode otp nya coba diisi pulsa nomer yg mau didaftarkan
5	Bagus dengan adanya apk ini jdi lebih mudah dan praktis

Tabel 1 menampilkan hasil seleksi data *scraping* yang diperoleh melalui *web scraping* dari basis data ulasan aplikasi *Mobile* JKN. Data awal dikumpulkan tanpa seleksi, kemudian difilter untuk mempertahankan kolom *score* dan *content*, memastikan fokus analisis sentimen tetap relevan. Sebanyak 10.000 ulasan dari 16 September 2018 hingga 10 November 2024 dipilih berdasarkan skor 1-5 dan diurutkan menurut relevansi guna meningkatkan akurasi dalam memahami opini pengguna.

Tabel 2. Contoh data seleksi distribusi panjang ulasan

Score	*Content*	*Content Length*
1	Saya mau daftar .. muncul No hp sudah digunakan.. Saya baru pertama kali pake aplikasi nya	90
4	Sangat membantu sekali.salam sehat dan sukses selalu kedepannya.ðŸ™□ðŸ™□ðŸ‘□ðŸ‘□ðŸ¤²ðŸ¤²	70
1	Ini aplikasi kenapa buruk sekali sistemnya. Verifikasi nomer telpon butuh waktu ratusan detik tapi tetap tidak ada. Bukannya memudahkan malah menyusahkan.	154
5	Buat yang gk masuk kode otp nya coba diisi pulsa nomer yg mau didaftarkan	73
5	Bagus dengan adanya apk ini jdi lebih mudah dan praktis	55

Tabel 2 menyajikan data hasil seleksi distribusi panjang ulasan yang awalnya dari 10.000 data, di mana sebanyak 8.738 ulasan memenuhi kriteria relevansi untuk analisis lebih lanjut. Pada dataset ini, ditambahkan kolom *content length* yang mengukur jumlah karakter dalam setiap ulasan guna menganalisis distribusi panjang teks. Berdasarkan hasil analisis, histogram distribusi panjang ulasan menunjukkan bahwa mayoritas ulasan berada dalam rentang 50–350 karakter. Ulasan dengan panjang kurang dari 50 karakter cenderung tidak memberikan informasi yang cukup untuk dianalisis secara efektif, sementara ulasan yang melebihi 350 karakter berpotensi mengandung informasi yang tidak terfokus dan kurang relevan. Oleh karena itu, untuk meningkatkan kualitas data dan memastikan analisis yang lebih akurat, ulasan di luar rentang tersebut dieliminasi.

4.2. *Pre-processing*

Tabel 3. Tahapan dan hasil *preprocessing*

Tahapan	**Hasil**
Ulasan Asli	Ini aplikasi kenapa buruk sekali sistemnya. Verifikasi nomer telpon butuh waktu ratusan detik tapi tetap tidak ada. Bukannya memudahkan malah menyusahkan.
Filtering & Cleansing	Ini aplikasi kenapa buruk sekali sistemnya Verifikasi nomer telpon butuh waktu ratusan detik tapi tetap tidak ada Bukannya memudahkan malah menyusahkan
Case Folding	ini aplikasi kenapa buruk sekali sistemnya verifikasi nomer telpon butuh waktu ratusan detik tapi tetap tidak ada bukannya memudahkan malah menyusahkan
Tokenizing	['ini', 'aplikasi', 'kenapa', 'buruk', 'sekali', 'sistemnya', 'verifikasi', 'nomer', 'telpon', 'butuh', 'waktu', 'ratusan', 'detik', 'tapi', 'tetap', 'tidak', 'ada', 'bukannya', 'memudahkan', 'malah', 'menyusahkan']
Normalisasi Teks	['ini', 'aplikasi', 'kenapa', 'buruk', 'sekali', 'sistemnya', 'verifikasi', 'nomor', 'telpon', 'butuh', 'waktu', 'ratusan', 'detik', 'tapi', 'tetap', 'tidak', 'ada', 'bukannya', 'memudahkan', 'malah', 'menyusahkan']

Tahapan	Hasil
Menghapus *Stopword*	['aplikasi', 'buruk', 'sistemnya', 'verifikasi', 'nomor', 'telpon', 'butuh', 'ratusan', 'detik', 'memudahkan', 'menyusahkan']
Stemming	['aplikasi', 'buruk', 'sistem', 'verifikasi', 'nomor', 'telpon', 'butuh', 'ratus', 'detik', 'mudah', 'susah']

Tabel 3 menggambarkan tahapan *pre-processing* yang diterapkan pada teks ulasan pengguna aplikasi. Proses ini penting untuk memastikan bahwa data teks yang digunakan dalam analisis sentimen bersih, terstruktur, dan siap untuk diproses lebih lanjut. Berikut adalah uraian setiap tahapan *preprocessing* yang dilakukan:

a. Tahap *filtering & cleansing* bertujuan untuk menghapus data yang tidak lengkap, duplikat pada kolom konten, serta elemen yang tidak relevan atau karakter yang dapat mengganggu analisis. Pada tahap *filtering*, data yang tidak memenuhi kriteria akan dihapus tanpa mengubah isi teks. Selanjutnya, pada tahap *cleansing*, dilakukan pemrosesan untuk memastikan teks sesuai dengan format standar, yang mencakup penggunaan huruf kecil dan besar (a-z), penghapusan spasi berlebih, serta koreksi spasi di awal dan akhir kalimat.
b. Tahap *case folding*, yang mengubah seluruh teks menjadi huruf kecil untuk menghilangkan perbedaan antara huruf kapital dan huruf kecil. Hal ini memastikan bahwa kata-kata yang sama tidak dianggap berbeda hanya karena perbedaan penggunaan huruf kapital.
c. Tahap *tokenizing*, pada tahap ini, teks dibagi menjadi token atau unit terkecil berupa kata-kata. *Tokenizing* memisahkan kalimat menjadi kata-kata individual yang lebih mudah dianalisis.
d. Tahap normalisasi teks ini bertujuan untuk menyamakan bentuk kata yang mungkin memiliki variasi ejaan, seperti "nomer" yang dinormalisasi menjadi "nomor". Hal ini penting untuk menghindari duplikasi entri kata yang sama namun dengan ejaan yang berbeda.
e. Tahap menghapus *stopword* dalam pengolahan teks untuk menghilangkan kata-kata yang dianggap tidak memberikan informasi penting dalam analisis sentimen, seperti kata sambung atau kata yang terlalu umum. Dalam contoh ini, kata-kata seperti "ini", "kenapa", "tapi", dan "ada" dihapus karena dianggap tidak relevan untuk menganalisis sentimen yang terkandung dalam teks ulasan.
f. Tahap *stemming* melibatkan pengubahan kata-kata menjadi bentuk dasarnya melalui penghapusan imbuhan. Misalnya, "memudahkan" diubah menjadi "mudah" dan "menyusahkan" menjadi "susah". Hal ini membantu dalam menyatukan variasi bentuk kata yang merujuk pada konsep yang sama.

4.3. *Transformation*

Tabel 4. Fitur-fitur dengan bobot TF-IDF tertinggi

Fitur	Nilai Rata-Rata TF-IDF	Nilai Chi-Square
aplikasi	0.077228	6.588881
daftar	0.062821	13.904371
nomor	0.060410	20.559000
masuk	0.044430	15.817497
verifikasi	0.043093	16.773988

Tabel 4 menampilkan fitur dengan bobot TF-IDF tertinggi, yang menunjukkan tingkat kepentingan kata dalam membedakan dokumen dalam korpus. TF-IDF dihitung berdasarkan frekuensi kata dalam dokumen (TF) dan *invers* frekuensi dokumen (IDF), sehingga kata yang sering muncul dalam dokumen tertentu tetapi jarang di dokumen lain memiliki bobot lebih tinggi. Misalnya, kata "aplikasi" (0.0772) dan "daftar" (0.0628) memiliki daya pembeda yang tinggi. Dengan demikian, fitur dengan bobot TF-IDF tertinggi berperan penting dalam analisis teks dan pemodelan dokumen.

Tabel 5. Fitur-fitur dengan nilai *chi-square* tertinggi

Fitur	Nilai rata-rata TF-IDF	Nilai C*hi-Square*
bantu	0.020177	171.031705
moga	0.006014	102.368692
mudah	0.030202	101.357154
sehat	0.007039	83.828751
alhamdulillah	0.001761	69.461905

Tabel 5 menampilkan fitur dengan nilai *chi-square* tertinggi, yang menunjukkan korelasi kuat dengan label sentimen. Misalnya, kata "bantu" (171.031) dan "mudah" (101.357) memiliki daya diskriminasi tinggi, dan kata "sehat" (83.828) juga berperan penting dalam membedakan sentimen. Fitur-fitur dengan nilai *chi-square* tinggi ini membantu meningkatkan akurasi model dalam klasifikasi sentimen.

4.4. *Data Mining*

Penelitian ini menggunakan *multinomial naive bayes* untuk klasifikasi sentimen karena efisien, mampu menangani data teks berdimensi tinggi, dan memiliki performa baik dalam analisis sentimen. Hasil dan pembahasan *data mining* ini akan dijelaskan sesuai dengan alur berikut.

a. Dimulai dengan tahap *input* data, yang bertujuan untuk memahami distribusi *dataset* yang digunakan dalam penelitian ini. Gambar 3 membantu memahami proporsi kelas dalam *dataset*, yang berpengaruh terhadap performa model klasifikasi sentimen.

```
jumlah data latih:
(6990, 500)
jumlah data uji:
(1748, 500)

Probabilitas Prior (DataFrame):
     Kelas  Jumlah Prior
0  negatif        5523.0
1   netral         510.0
2  positif         957.0
```

Gambar 3. Distribusi data

Gambar 3 menunjukkan distribusi dataset yang digunakan dalam penelitian ini, terdiri dari 6.990 sampel data latih dengan 5.523 sampel berlabel negatif, 510 sampel netral, dan 957 sampel positif. Selain itu, terdapat 1.748 sampel data uji yang digunakan untuk mengevaluasi performa model.

b. Langkah selanjutnya adalah perhitungan probabilitas *prior*, yang dilakukan untuk menentukan distribusi awal setiap kelas sentimen berdasarkan data latih menggunakan rumus:

$$P(c) = \frac{N_c}{N} \quad (5)$$

$$P(negatif) = \frac{5523}{6990} = 0{,}0790128755 \quad (6)$$

$$P(netral) = \frac{510}{6990} = 0{,}0729613734 \quad (7)$$

$$P(positif) = \frac{957}{6990} = 0{,}136909871 \quad (8)$$

c. Selanjutnya, dilakukan perhitungan probabilitas *likelihood* kata, yaitu menghitung peluang kemunculan setiap kata dalam masing-masing kelas sentimen.

Tabel 6. Beberapa contoh probabilitas *likelihood*

Kata	P(kata| negatif)	P(kata| netral)	P(kata| positif)
cepat	0,0018	0,003	0,0099
mudah	0,0086	0,0062	0,0301
layan	0,0044	0,0042	0,0142
guna	0,0051	0,0031	0,0062

Tabel 6 menyatakan bahwa nilai probabilitas *likelihood* digunakan dalam perhitungan selanjutnya untuk menentukan kecenderungan sentimen dari suatu dokumen berdasarkan kata-kata yang terkandung di dalamnya.

d. Tahap terakhir adalah prediksi sentimen dokumen, yang dilakukan dengan menghitung probabilitas *posterior* berdasarkan Teorema Bayes menggunakan rumus:

$$P(c|d) = P(c) \times \prod_{i=1}^{n} P(t_i|c) \quad (9)$$

Sebagai ilustrasi, jika sebuah dokumen mengandung kata "cepat mudah mudah layan guna", maka perhitungan probabilitas *posterior* untuk masing-masing kelas sebagai berikut:

$$P(negatif|"cepat\ mudah\ mudah\ layan\ guna") = P(negatif) \times P(cepat|negatif) \times P(mudah|negatif)^2 \times P(layan|negatif) \times P(guna|negatif) = 0{,}790128755 \times 0.0018 \times 0.0086^2 \times 0{,}0044 \times 0{,}0051 = 2.360424574498162 \times 10^{-12} \quad (10)$$

$$P(netral|"cepat\ mudah\ mudah\ layan\ guna") = P(netral) \times P(cepat|netral) \times P(mudah|netral)^2 \times P(layan|netral) \times P(guna|netral) = 0{,}0729613734 \times 0.0030 \times 0.0062^2 \times 0{,}0042 \times 0{,}0031 = 2.055655711348344 \times 10^{-13} \quad (11)$$

$$P(positif|"cepat\ mudah\ mudah\ layan\ guna") = P(positif) \times P(cepat|positif) \times P(mudah|positif)^2 \times P(layan|positif) \times P(guna|positif) = 0{,}136909871 \times 0.0099 \times 0.0301^2 \times 0{,}0142 \times 0{,}0062 = 1{,}0811426 \times 10^{-10} \quad (12)$$

Untuk membandingkan probabilitas *posterior* dalam bentuk persentase, kita perlu menghitung proporsi setiap probabilitas terhadap total probabilitas dari semua kelas. Berikut perhitungannya:

$$P\left(negatif \middle| \begin{matrix} "cepat\ mudah\ mudah \\ layan\ guna" \end{matrix}\right) = \frac{2{,}3\ldots\times10^{-12}}{2{,}3\ldots\times10^{-12}+2{,}0\ldots\times10^{-13}+1{,}0\ldots\times10^{-10}} \approx 2{,}13\% \quad (13)$$

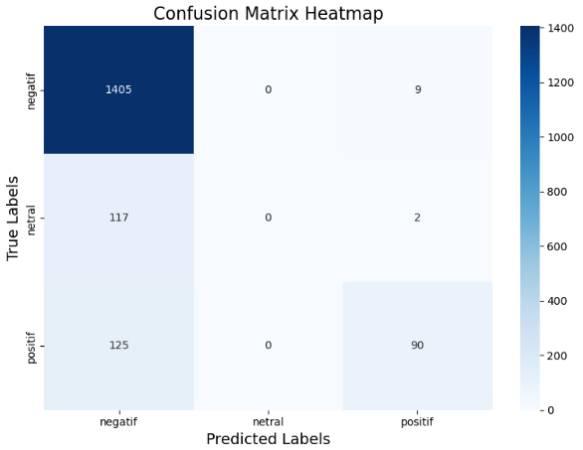
$$P\left(netral \middle| \begin{matrix} "cepat\ mudah\ mudah \\ layan\ guna" \end{matrix}\right) = \frac{2{,}0\ldots\times10^{-13}}{2{,}3\ldots\times10^{-12}+2{,}0\ldots\times10^{-13}+1{,}0\ldots\times10^{-10}} \approx 0{,}18\% \quad (14)$$

$$P\left(positif \middle| \begin{matrix} "cepat\ mudah\ mudah \\ layan\ guna" \end{matrix}\right) = \frac{1{,}0\ldots\times10^{-10}}{2{,}3\ldots\times10^{-12}+2{,}0\ldots\times10^{-13}+1{,}0\ldots\times10^{-10}} \approx 97{,}68\% \quad (15)$$

Dalam simulasi perhitungan, dokumen yang mengandung kata "cepat mudah mudah layan guna", menghasilkan nilai 2,13% untuk negatif, 0,18% untuk netral, dan 97,68% untuk positif. Karena probabilitas tertinggi berada pada kelas positif, maka dokumen tersebut diklasifikasikan sebagai bersentimen positif.

4.5. *Interpretation & Evaluation*

Tahap *Interpretation & Evaluation* bertujuan untuk menilai kinerja model *Multinomial Naïve Bayes* dalam mengklasifikasikan sentimen ulasan aplikasi *Mobile* JKN. Evaluasi dilakukan menggunakan *confusion matrix*, yang memberikan gambaran mengenai akurasi klasifikasi dengan membandingkan hasil prediksi model dengan label sebenarnya. Matriks ini menunjukkan jumlah prediksi yang benar dan salah untuk setiap kategori sentimen, yaitu 'negatif', 'netral', dan 'positif', sekaligus mengidentifikasi kesalahan prediksi.



Gambar 4. Hasil *confusion matrix*

Pada *confusion matrix* (gambar 4), angka diagonal utama (1405, 0, 90) menunjukkan jumlah data yang diklasifikasi dengan benar, sementara angka di luar diagonal mencerminkan kesalahan prediksi. Rinciannya, 9 ulasan negatif salah diklasifikasikan sebagai positif, 117 ulasan netral diklasifikasikan sebagai negatif, dan 2 ulasan netral sebagai positif. Selain itu, 125 ulasan positif keliru diklasifikasikan sebagai negatif.

Metrik evaluasi digunakan untuk mengukur kinerja model berdasarkan hasil *confusion matrix*. Berikut ini adalah perhitungan beberapa metrik utama:

a. Akurasi
Diketahui:
TP = 1405 + 0 + 90 = 1.495
Total Prediksi = 1405 + 0 + 9 + 117 + 0 + 2 + 125 + 0 + 90 = 1.748
Hasil Perhitungan:

$$Accuracy = \frac{\text{TP}}{\text{Total Prediksi}} = \frac{1495}{1748} \approx 0{,}855 \quad (16)$$

Model *multinomial naïve bayes* memiliki akurasi 85,5%, yang berarti dari 1.748 data, sebanyak 1.495 diklasifikasikan dengan benar.

b. Kelas Negatif
Diketahui:
TP = 1.405
FP = 117 + 125 = 242
FN = 0 + 9 = 9
Hasil Perhitungan:

$$Precision = \frac{\text{TP}}{\text{TP+FP}} = \frac{1405}{1405+242} \approx 0{,}853 \quad (17)$$

$$Recal = \frac{\text{TP}}{\text{TP+FN}} = \frac{1405}{1405+9} \approx 0{,}994 \quad (18)$$

Model memiliki presisi 85,3%, artinya dari semua prediksi negatif, 85,3% benar-benar negatif. *Recall* 99,4% menunjukkan hampir semua ulasan negatif terdeteksi dengan baik.

c. Kelas Netral
Diketahui:
TP = 0
FP = 0 + 0 = 0
FN = 117 + 2 = 119
Hasil Perhitungan:

$$Precision = \frac{\text{TP}}{\text{TP+FP}} = \frac{0}{0+0} = \infty \text{ (dianggap 0)} \quad (19)$$

$$Recall = \frac{\text{TP}}{\text{TP+FN}} = \frac{0}{0+(117+2)} = 0 \quad (20)$$

Model tidak pernah memprediksi ulasan sebagai netral, menyebabkan *precision* tidak tentu (dianggap 0%) dan *recall* 0%.

d. Kelas Positif
Diketahui:
TP = 90
FP = 9 + 2 = 11
FN = 0 + 125 = 125
Hasil Perhitungan:

$$Precision = \frac{\text{TP}}{\text{TP+FP}} = \frac{90}{90+11} \approx 0{,}891 \quad (21)$$

$$Recall = \frac{\text{TP}}{\text{TP+FN}} = \frac{90}{90+125} \approx 0{,}419 \quad (22)$$

Precision 89,1% menunjukkan sebagian besar prediksi positif benar, tetapi *recall* hanya 41,9%, menandakan banyak ulasan positif tidak terdeteksi.

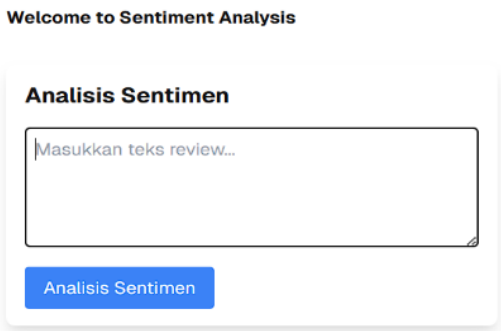
Classification Report:

	precision	recall	f1-score	support
negatif	0.85	0.99	0.92	1414
netral	0.00	0.00	0.00	119
positif	0.89	0.42	0.57	215
accuracy			0.86	1748
macro avg	0.58	0.47	0.50	1748
weighted avg	0.80	0.86	0.81	1748

Gambar 5. *Output* evaluasi model *confusion matrix*

Model dengan kombinasi parameter data uji (20%), TF-IDF (1000 fitur), dan Chi-Square (500 fitur) mencapai akurasi 86% dengan performa tinggi dalam mengidentifikasi ulasan negatif (*recall* tinggi) dan ketepatan prediksi untuk ulasan negatif serta positif (*precision* tinggi). Namun, model gagal mengenali ulasan netral dan kurang efektif dalam menemukan semua ulasan positif (*recall* rendah). Keterbatasan ini diduga akibat representasi data yang kurang optimal dalam membedakan kelas netral dari yang lain. Model cenderung bias terhadap sentimen ekstrem, sehingga diperlukan perbaikan dalam pemrosesan data dan pemilihan fitur agar dapat mengenali kelas netral dengan lebih akurat.

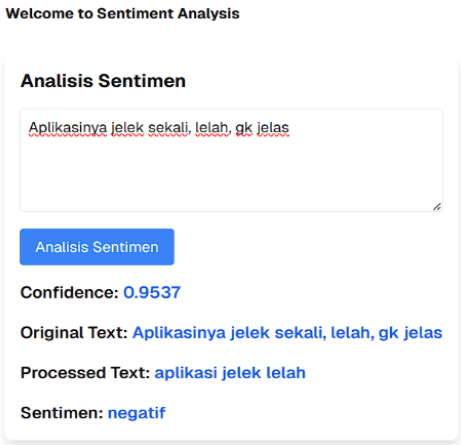
4.6. Penerapan Aplikasi



Gambar 6. Tampilan *web app*

Gambar 6 menunjukkan antarmuka pengguna aplikasi *web* "Analisis Sentimen App". Pengguna memasukkan teks ulasan atau opini ke dalam kotak teks, lalu menekan tombol "Analisis Sentimen" untuk memproses teks tersebut menggunakan model *machine learning* dan menentukan sentimennya.

Informasi yang ditampilkan pada aplikasi "Analisis Sentimen App" mencakup beberapa komponen utama. Pertama, *confidence score* atau tingkat kepercayaan model terhadap hasil klasifikasi sentimen. Kedua, teks asli (*original text*) yang dimasukkan pengguna sebelum diproses. Ketiga, teks yang telah melalui tahap *preprocessing*. Keempat, kategori sentimen hasil klasifikasi.



Gambar 7. Hasil analisis sentimen pada *web app*

Gambar 7 menunjukkan hasil analisis sentimen terhadap ulasan "Aplikasinya jelek sekali, lelah, gk jelas", yang diklasifikasikan sebagai negatif dengan nilai *confidence* sebesar 95,37%. Setelah melalui tahap *preprocessing*, teks disederhanakan menjadi "aplikasi jelek lelah", sehingga memudahkan model dalam mengidentifikasi sentimen secara lebih efektif.

5. KESIMPULAN DAN SARAN

Model *multinomial naïve bayes* mencapai akurasi 86% dengan performa terbaik pada ulasan negatif (*f1-score* 92%), tetapi kurang efektif dalam mengenali ulasan netral (*f1-score* = 0) dan positif (*f1-score* 57%), yang menyebabkan nilai *macro average* jauh lebih rendah dari *weighted average*, mencerminkan ketidakseimbangan akibat dominasi kelas negatif. Untuk meningkatkan performa, perlu dilakukan perbaikan representasi kelas netral melalui analisis mendalam, augmentasi data, atau penyeimbangan kelas. Peningkatan *preprocessing*, seperti normalisasi teks dan penghapusan *noise*, serta optimasi fitur dengan TF-IDF dan *chi-square* atau pendekatan canggih seperti Word2Vec dan BERT dapat membantu. Eksplorasi algoritma lain seperti SVM, LSTM, dan *ensemble learning*, serta validasi model dengan *cross-validation* dan *hyperparameter tuning* diharapkan dapat meningkatkan keseimbangan dan akurasi klasifikasi.

DAFTAR PUSTAKA

[1] K. Kesehatan, “Transformasi kesehatan digital: Aksesibilitas dan afordabilitas,” *Kementerian Kesehatan Republik Indonesia*, 2024. https://rc.kemkes.go.id/transformasi-kesehatan-digital-aksesibilitas-dan-afordabilitas-0b87bf (accessed Jan. 13, 2025).

[2] O. B. Kusumawardhani, A. Octaviana, and Y. M. Supitra, “Efektivitas Mobile JKN Bagi Masyarakat : Literature Review,” *Pros. Semin. Inf. Kesehat. Nas.*, pp. 64–69, 2022, [Online]. Available: https://ojs.udb.ac.id/index.php/sikenas/article/view/1665

[3] S. Roiqoh, B. Zaman, and K. Kartono, “Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1582, 2023, doi: 10.30865/mib.v7i3.6194.

[4] E. W. Sholeha, S. Yunita, R. Hammad, V. C. Hardita, and K. Kaharuddin, “Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 203–208, 2022, doi: 10.35746/jtim.v3i4.178.

[5] S. Narmansyah, Indar, S. Rahmadani, M. A. Arifin, and R. M Thaha, “Analisis Pemanfaatan Sistem Informasi JKN Mobile Di Kota Makassar,” *Sehat Rakyat J. Kesehat. Masy.*, vol. 1, no. 3, pp. 196–204, 2022, doi: 10.54259/sehatrakyat.v1i3.1082.

[6] amar Nanda Syarif, N. Pambudiyatno, W. Utomo, J. I. Jemur Andayani No, and K. Siwalankerto Kec Wonocolo, “Rancangan Sistem Presensi Dan Rekapitulasi Jurnal Kegiatan Ojt Menggunakan Visual Studio Code Berbasis Web Di Airnav Cabang Matsc,” *Pros. Semin. Nas. Inov. Teknol. Penerbangan Tahun*, p. 2023, 2023.

[7] R. F. Muharram and A. Suryadi, “Implementasi Artificial Intelligence Untuk Deteksi Masker Secara Realtime Dengan Tensorflow Dan SSD Mobilenet Berbasis Pyton,” *JRKT (Jurnal Rekayasa Komputasi Ter.*, vol. 1, no. 03, pp. 281–290, 2021, doi: 10.30998/jrkt.v1i03.5832.

[8] P. Apriyani, A. R. Dikananda, and I. Ali, “Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi,” *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.

[9] Y. Kartika, K. Komariah, A. Surip, R. Saputra, and I. Ali, “Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Rotan,” *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 1, pp. 28–34, 2022, doi: 10.32485/kopertip.v4i1.112.

[10] E. Hokijuliandy, H. Napitupulu, and Firdaniza, “Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia’s

National Health Insurance Mobile Application," *Mathematics*, vol. 11, no. 17, pp. 0–21, 2023, doi: 10.3390/math11173765.

[11] M. R. Ramadhan, E. Budianita, I. Iskandar, and M. Affandes, "Analisis Sentimen Masyarakat Mengenai Aplikasi Shopeefood Menggunakan Chi Square dan Metode Naive Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 1170–1178, 2023, doi: 10.30865/klik.v3i6.927.

[12] C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Portal Data*, vol. 2, no. 3, pp. 1–12, 2022, [Online]. Available: http://portaldata.org/index.php/portaldata/article/view/107

[13] T. Ramadhani, P. Hermawan, and A. R. Dzikrillah, "Penerapan Metode Naïve Bayes untuk Analisis Sentimen pada Ulasan Pengguna Aplikasi ChatGPT di Google Play Store," *Technol. Sci.*, vol. 6, no. 1, pp. 430–439, 2024, doi: 10.47065/bits.v6i1.5400.

[14] A. Fahrezi and A. Solichin, "ANALISIS SENTIMEN ULASAN APLIKASI MOBILE JKN PADA PLAY STORE MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES," *Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 3, no. 2, pp. 344–353, 2024.

[15] R. AL Anshari, S. Alam, and M. Hafid T, "Komparasi Payment Digital Untuk Analisis Sentimen Berdasarkan Ulasan Di Google Playstore Menggunakan Metode Support Vector Machine," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 118–128, 2023, doi: 10.55123/storage.v2i3.2337.

[16] D. F. Sari, A. Kusjani, D. Kurniawati, and I. Setiawan, "PENCARIAN DATA QUICK COUNT PILPRES DENGAN TEKNIK WEB SCRAPING," *J. Innov. Res. Knowl.*, vol. 3, no. 5, pp. 1025–1034, 2023.