Optimasi Nilai K pada Algoritma k-NN untuk Analisis Sentimen Ulasan Aplikasi Gopay Menggunakan Python

Anas Nasrulloh[1)], Tubagus Toifur[2)], Aolia Ikhwanudin[3)], Muhammad Yusuf[4)], Ibnu Mas'ud[5)]

[1)3)]Sistem Informasi, Ilmu Komputer, Institut Teknologi Tangerang Selatan
[2)5)]Teknologi Informasi, Ilmu Komputer, Institut Teknologi Tangerang Selatan
[4)]Informatika, Ilmu Komputer, Institut Teknologi Tangerang Selatan
Serpong Utara, Tangerang Selatan
[1)]anas@itts.ac.id
[2)]tubagus@itts.ac.id
[3)]aikwanudin@itts.ac.id
[4)]yusuf@itts.ac.id
[5)]ibnu@itts.ac.id

Abstrak

Pada penelitian ini ditemukan bahwa memilih nilai k yang terlalu besar dapat mengurangi performa model, sementara nilai k yang lebih kecil memberikan hasil yang lebih optimal dalam analisis sentimen ulasan aplikasi. Penelitian ini berfokus pada optimasi nilai k dalam algoritma *k-Nearest Neighbors* (k-NN) untuk analisis sentimen ulasan aplikasi Gopay yang diperoleh melalui data scraping dari Google Play Store. Peneliti mengumpulkan dan menganalisis data ulasan untuk menilai akurasi algoritma k-NN pada berbagai nilai k. Nilai k yang diuji meliputi 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, dan 49. Hasil analisis menunjukkan bahwa akurasi algoritma k-NN mencapai puncaknya pada nilai k = 3 dengan akurasi 0.898, sedangkan pada nilai k yang lebih besar, akurasi cenderung menurun secara bertahap hingga mencapai 0.870 pada k=47 dan k=49.

Kata kunci: Algoritma *k-Nearest Neighbors* (k-NN), Analisis Sentimen, Optimasi Nilai K

Abstract

This study found that choosing a k value that is too large can reduce model performance, while a smaller k value provides more optimal results in the analysis of app review sentiment. This study focuses on optimizing the k value in the k-Nearest Neighbors (k-NN) algorithm for sentiment analysis of Gopay app reviews obtained through data scraping from the Google Play Store. Researchers collected and analyzed review data to assess the accuracy of the k-NN algorithm at various k values. The tested k values include 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, and 49. The analysis results show that the accuracy of the k-NN algorithm peaks at a value of k = 3 with an accuracy of 0.898, while at larger k values, the accuracy tends to decrease gradually until it reaches 0.870 at k = 47 and k = 49.

Keywords: *K Value Optimization, k-Nearest Neighbors (k-NN) Algorithm, Sentiment Analysis*

1. PENDAHULUAN

Aplikasi Gopay memungkinkan pengguna untuk melakukan transaksi keuangan secara efisien melalui perangkat mobile mereka. Dengan bertambahnya jumlah pengguna aplikasi ini, ulasan yang diberikan oleh pengguna di platform seperti Google Play Store menjadi sumber informasi yang sangat berharga. Ulasan ini tidak hanya mencerminkan pengalaman pengguna tetapi juga dapat memberikan wawasan penting tentang kualitas dan performa aplikasi. Dalam hal ini,

DOI: 10.34128/jsi.v12i1.1123 *Received: September 4th, 2024*
Accepted: April 12th, 2026

analisis sentimen terhadap ulasan aplikasi dapat memberikan pemahaman yang lebih baik tentang persepsi pengguna, serta mengidentifikasi kekuatan dan kelemahan dari aplikasi tersebut.

Analisis sentimen adalah metode yang digunakan untuk menilai sikap atau perasaan seseorang terhadap suatu produk atau layanan berdasarkan teks yang ditulis, seperti ulasan pengguna. Tujuannya adalah untuk mengklasifikasikan teks ke dalam kategori sentimen yang telah ditentukan, seperti positif, negatif, atau netral. Dalam era informasi yang melimpah ini, analisis sentimen telah menjadi alat penting dalam mengolah data besar dan mendapatkan wawasan yang berarti. Dengan adanya teknik analisis sentimen, perusahaan dapat mengevaluasi umpan balik pengguna secara lebih sistematis dan membuat keputusan yang berbasis data.

Penelitian sebelumnya menunjukkan bahwa kombinasi metode SVM, Grid Search, dan N-Gram efektif meningkatkan akurasi klasifikasi sentimen pada ulasan game mobile. Metode ini meningkatkan ketepatan hasil dan menunjukkan pentingnya pemilihan fitur serta pengaturan parameter model. Penulis menyarankan eksperimen dengan metode lain, seperti deep learning, untuk hasil yang lebih baik [1]. Penelitian sebelumnya menunjukkan bahwa menggabungkan algoritma SVM dengan pemilihan fitur menggunakan Gain Ratio dapat meningkatkan analisis sentimen. Pemilihan fitur yang tepat menyederhanakan data, mempercepat pelatihan model, dan meningkatkan akurasi klasifikasi. Gain Ratio terbukti efektif untuk memilih fitur relevan dan mengoptimalkan model SVM dalam klasifikasi teks [2]. Penelitian sebelumnya menunjukkan bahwa menggabungkan algoritma Naive Bayes dengan teknik Chi-Square untuk memilih fitur efektif meningkatkan akurasi analisis sentimen pada komentar produk Xiaomi SU7 di YouTube. Metode ini menghasilkan klasifikasi yang lebih akurat dibandingkan metode lain dan memberikan kontribusi penting dalam analisis sentimen di platform besar seperti YouTube [3]. Penelitian sebelumnya menunjukkan bahwa menggunakan Particle Swarm Optimization (PSO) untuk meningkatkan kinerja Decision Tree adalah cara yang efektif untuk meningkatkan akurasi dalam analisis sentimen pada ulasan game GTA V Roleplay. Dengan mengoptimalkan parameter model Decision Tree menggunakan PSO, penelitian ini berhasil menunjukkan hasil yang lebih baik dalam mengklasifikasikan sentimen ulasan dibandingkan dengan menggunakan Decision Tree biasa [4]. Penelitian sebelumnya menunjukkan bahwa menggunakan Algoritma Genetika (GA) untuk mengoptimalkan parameter Support Vector Machine (SVM) sangat efektif dalam analisis sentimen pada data teks dari media sosial Instagram. Dengan optimasi ini, model SVM dapat memberikan hasil yang lebih akurat dan cepat dalam mengklasifikasikan sentimen pada komentar atau postingan pengguna [5]. Penelitian sebelumnya menunjukkan bahwa menggunakan Particle Swarm Optimization (PSO) untuk mengoptimalkan parameter Support Vector Machine (SVM) sangat efektif dalam meningkatkan akurasi analisis sentimen terkait wacana pindah ibu kota Indonesia. Dengan optimasi ini, SVM dapat memberikan hasil klasifikasi yang lebih baik dan efisien, sehingga menghasilkan wawasan yang lebih akurat tentang pandangan masyarakat terhadap isu ini [6].

Salah satu algoritma pembelajaran mesin yang sering digunakan dalam analisis sentimen adalah k-Nearest Neighbors (k-NN). Algoritma k-NN adalah metode yang sederhana namun efektif dalam tugas klasifikasi dan regresi. Konsep dasar dari algoritma ini adalah mengklasifikasikan data berdasarkan kedekatannya dengan data lain yang sudah diketahui kategorinya. Dalam proses klasifikasi, k-NN mencari k tetangga terdekat dari data yang akan diklasifikasikan dan memberikan prediksi berdasarkan mayoritas kategori dari tetangga tersebut. Keberhasilan algoritma k-NN sangat bergantung pada pemilihan nilai k, yaitu jumlah tetangga yang dipertimbangkan dalam proses klasifikasi. Nilai k yang terlalu kecil dapat menyebabkan model menjadi terlalu sensitif terhadap noise dalam data, sedangkan nilai k yang terlalu besar dapat mengakibatkan model terlalu generalisasi dan kehilangan detail penting. Oleh karena itu, optimasi nilai k adalah langkah krusial dalam meningkatkan akurasi dan efektivitas algoritma k-NN.

Dalam penelitian ini, kami melakukan beberapa langkah metodologis yang meliputi pengumpulan data, preprocessing data, pelatihan model, pengujian, dan evaluasi hasil. Pengumpulan data dilakukan dengan teknik scraping untuk mendapatkan ulasan aplikasi dari Google Play Store. Data yang diperoleh kemudian dibersihkan dan diproses untuk menghilangkan noise serta format yang tidak konsisten. Langkah ini termasuk tokenisasi, penghapusan stop words, dan stemming untuk mempersiapkan data bagi pelatihan model. Setelah data siap, model k-NN dilatih dengan berbagai

nilai k untuk mengklasifikasikan ulasan ke dalam kategori sentimen. Model yang telah dilatih diuji dengan data yang belum pernah dilihat sebelumnya untuk mengevaluasi akurasi klasifikasinya.

2. TINJAUAN PUSTAKA

Beberapa studi literature dilakukan peneliti terhadap penelitian sebelumnya yang relevan, diantaranya sebagai berikut:

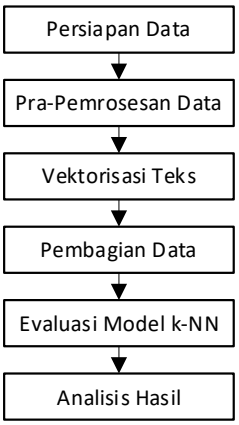
- Penelitian ini menunjukkan bahwa kombinasi metode SVM, Grid Search, dan N-Gram efektif meningkatkan akurasi klasifikasi sentimen pada ulasan game mobile. Metode ini meningkatkan ketepatan hasil dan menunjukkan pentingnya pemilihan fitur serta pengaturan parameter model. Penulis menyarankan eksperimen dengan metode lain, seperti deep learning, untuk hasil yang lebih baik [1]
- Penelitian ini menunjukkan bahwa menggabungkan algoritma SVM dengan pemilihan fitur menggunakan Gain Ratio dapat meningkatkan analisis sentimen. Pemilihan fitur yang tepat menyederhanakan data, mempercepat pelatihan model, dan meningkatkan akurasi klasifikasi. Gain Ratio terbukti efektif untuk memilih fitur relevan dan mengoptimalkan model SVM dalam klasifikasi teks [2].
- Penelitian ini menunjukkan bahwa menggabungkan algoritma Naive Bayes dengan teknik Chi-Square untuk memilih fitur efektif meningkatkan akurasi analisis sentimen pada komentar produk Xiaomi SU7 di YouTube. Metode ini menghasilkan klasifikasi yang lebih akurat dibandingkan metode lain dan memberikan kontribusi penting dalam analisis sentimen di platform besar seperti YouTube [3].
- Penelitian ini menunjukkan bahwa menggunakan Particle Swarm Optimization (PSO) untuk meningkatkan kinerja Decision Tree adalah cara yang efektif untuk meningkatkan akurasi dalam analisis sentimen pada ulasan game GTA V Roleplay. Dengan mengoptimalkan parameter model Decision Tree menggunakan PSO, penelitian ini berhasil menunjukkan hasil yang lebih baik dalam mengklasifikasikan sentimen ulasan dibandingkan dengan menggunakan Decision Tree biasa [4].
- Penelitian ini menunjukkan bahwa menggunakan Algoritma Genetika (GA) untuk mengoptimalkan parameter Support Vector Machine (SVM) sangat efektif dalam analisis sentimen pada data teks dari media sosial Instagram. Dengan optimasi ini, model SVM dapat memberikan hasil yang lebih akurat dan cepat dalam mengklasifikasikan sentimen pada komentar atau postingan pengguna [5].
- Penelitian ini menunjukkan bahwa menggunakan Particle Swarm Optimization (PSO) untuk mengoptimalkan parameter Support Vector Machine (SVM) sangat efektif dalam meningkatkan akurasi analisis sentimen terkait wacana pindah ibu kota Indonesia. Dengan optimasi ini, SVM dapat memberikan hasil klasifikasi yang lebih baik dan efisien, sehingga menghasilkan wawasan yang lebih akurat tentang pandangan masyarakat terhadap isu ini [6].

3. METODE PENELITIAN

Data dikumpulkan dari ulasan pengguna pada aplikasi gopay di Google Play Store pada tanggal 30 Juli 2024 sebanyak 67568 ulasan yang mana ulasan tersebut dari tanggal 22 Februari 2024 sampai dengan tanggal 28 Juli 2024, struktur data yang diperoleh dari proses scraping memiliki kolom-kolom berikut:

- username, time
- rating: rating numerik yang diberikan oleh pengguna, biasanya dalam skala 1 hingga 5. Rating ini digunakan untuk menentukan sentimen serta memahami distribusi rating aplikasi.
- content: merupakan teks utama dari ulasan yang menjadi fokus utama dalam analisis sentimen dan pemrosesan teks. Teks ini digunakan untuk menilai sentimen—apakah positif, negatif, atau netral—serta untuk mengekstrak informasi penting lainnya.

Dalam penelitian ini, menggunakan langkah-langkah seperti terlihat pada Gambar 1.



Gambar 1. Langkah-langkah Penelitian

Dari Gambar 1 dapat dijelaskan sebagai berikut:

3.1 Persiapan Data

Pada tahap Persiapan Data, langkah pertama adalah Mengimpor Data, di mana dataset ulasan aplikasi Gopay diambil dari file CSV menggunakan library pandas. Proses ini dilakukan dengan menggunakan perintah data = pd.read_csv('Data ulasan aplikasi gopay.csv'), yang memungkinkan kita untuk memuat data ke dalam DataFrame sehingga dapat dikelola dan dianalisis lebih lanjut dalam lingkungan Python. Setelah data berhasil diimpor, langkah berikutnya adalah Pra-pemrosesan Data. Pada tahap ini, kita menangani nilai yang hilang di kolom content dengan menggantinya dengan string kosong menggunakan metode fillna(''). Hal ini penting untuk menghindari masalah yang mungkin timbul akibat adanya nilai kosong dalam data yang dapat mengganggu analisis lebih lanjut. Selain itu, kita juga memastikan bahwa kolom content memiliki tipe data yang konsisten dengan mengubahnya menjadi string menggunakan astype(str). Ini menjamin bahwa semua entri dalam kolom tersebut diperlakukan sebagai teks, sehingga memudahkan proses pembersihan dan pemrosesan data selanjutnya.

3.2 Pra-pemrosesan Data

Pada tahap Pra-pemrosesan Teks, langkah pertama adalah Mendefinisikan Fungsi Pra-pemrosesan yang bertujuan untuk membersihkan dan menyiapkan teks agar siap untuk analisis lebih lanjut. Fungsi ini melakukan beberapa proses penting: pertama, teks diubah menjadi huruf kecil untuk menghindari perbedaan antara huruf kapital dan kecil yang tidak perlu, sehingga membuat analisis lebih konsisten. Selanjutnya, karakter non-alfanumerik, seperti tanda baca dan simbol, dihapus dari teks, kecuali spasi, untuk menghilangkan elemen-elemen yang tidak relevan dan memfokuskan analisis hanya pada informasi penting yang ada dalam kata-kata. Fungsi ini kemudian melanjutkan dengan Tokenisasi Teks, yaitu memecah teks menjadi kata-kata individu, dan menghapus kata-kata berhenti (stop words) yang sering kali tidak memberikan kontribusi berarti terhadap analisis, seperti kata sambung atau kata penghubung.

Setelah fungsi pra-pemrosesan didefinisikan, langkah berikutnya adalah Menerapkan Fungsi Pra-pemrosesan pada data yang ada. Dalam konteks ini, fungsi preprocess_text diterapkan pada kolom content dalam DataFrame untuk membersihkan dan memformat teks ulasan. Hasil dari proses ini disimpan dalam kolom baru yang disebut cleaned_content, yang berisi teks yang telah dipra-pemrosesan dan siap untuk digunakan dalam analisis lebih lanjut atau sebagai input untuk model machine learning. Dengan menerapkan fungsi ini, kita memastikan bahwa teks yang digunakan dalam analisis adalah bersih, konsisten, dan relevan, yang dapat meningkatkan akurasi dan efektivitas model yang akan digunakan.

3.3 Vektorisasi Teks

Pada tahap Vektorisasi Teks, langkah pertama adalah Inisialisasi Vektorisasi, di mana CountVectorizer digunakan untuk mengonversi teks yang telah dipra-pemrosesan menjadi fitur numerik. CountVectorizer adalah alat dari library scikit-learn yang mengubah teks mentah menjadi representasi numerik dengan menghitung frekuensi kemunculan setiap kata dalam dokumen. Proses ini memungkinkan teks, yang pada dasarnya adalah data tidak terstruktur, diubah menjadi bentuk yang dapat diproses oleh algoritma machine learning.

Setelah inisialisasi, langkah berikutnya adalah Transformasi Data. Pada tahap ini, CountVectorizer diterapkan pada kolom cleaned_content, yang berisi teks yang telah dibersihkan dan dipra-pemrosesan. Hasil dari aplikasi CountVectorizer adalah matriks fitur X, yang menyajikan representasi numerik dari setiap dokumen dalam bentuk frekuensi kata. Selain itu, target variabel y diambil dari kolom sentiment, yang memberikan label atau kategori sentimen untuk masing-masing ulasan. Dengan demikian, proses ini menghasilkan dua elemen penting: matriks fitur X, yang digunakan sebagai input untuk model machine learning, dan y, yang berfungsi sebagai output atau label target. Transformasi ini memungkinkan model untuk memproses dan menganalisis data teks secara efektif dengan memanfaatkan representasi numerik yang terstruktur.

3.4 Pembagian Data

Pada tahap Pembagian Data, langkah utama adalah Pisahkan Data untuk Pelatihan dan Pengujian. Proses ini dilakukan dengan menggunakan fungsi train_test_split dari library scikit-learn, yang membagi dataset menjadi dua subset: data pelatihan dan data pengujian. Dalam langkah ini, dataset dipecah dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Data pelatihan digunakan untuk melatih model machine learning, yaitu proses di mana model belajar mengenali pola dan hubungan dalam data. Sementara itu, data pengujian disisihkan untuk mengevaluasi kinerja model setelah proses pelatihan selesai. Pembagian ini penting karena memastikan bahwa model dievaluasi pada data yang tidak terlihat sebelumnya, yang memberikan gambaran yang lebih akurat tentang seberapa baik model akan bekerja pada data baru dan tidak terduga di dunia nyata. Dengan melakukan pembagian data secara proporsional, kita dapat menjaga keseimbangan antara pelatihan model dan pengujian kinerjanya untuk mendapatkan hasil yang andal dan representatif.

3.5 Evaluasi Model k-NN

Pada tahap Evaluasi Model K-NN, langkah pertama adalah Daftar Nilai K, di mana kita menentukan berbagai nilai untuk parameter n_neighbors dari model K-Nearest Neighbors (K-NN) yang akan diuji. Parameter n_neighbors menentukan jumlah tetangga terdekat yang akan dipertimbangkan oleh model untuk membuat keputusan klasifikasi. Dengan menyusun daftar nilai k, kita dapat mengeksplorasi bagaimana perubahan dalam parameter ini mempengaruhi kinerja model.

Langkah berikutnya adalah Latih dan Uji Model untuk Setiap Nilai K. Untuk setiap nilai k dalam daftar, kita membuat sebuah model KNeighborsClassifier dengan n_neighbors=k. Model ini kemudian dilatih menggunakan data pelatihan (X_train, y_train) untuk mempelajari pola dari data. Setelah pelatihan, model diuji pada data pengujian (X_test) untuk memprediksi hasil berdasarkan informasi yang telah dipelajari. Akurasi dari prediksi model dihitung untuk setiap nilai k, memberikan ukuran seberapa baik model dalam mengklasifikasikan data pengujian. Hasil akurasi ini kemudian disimpan untuk analisis lebih lanjut.

Langkah terakhir adalah Simpan dan Tampilkan Hasil. Hasil akurasi yang diperoleh untuk setiap nilai k disimpan dalam sebuah DataFrame yang disebut results_df. DataFrame ini memungkinkan kita untuk menyimpan dan mengelola hasil evaluasi secara terstruktur. Setelah hasil disimpan, tabel hasil akurasi ditampilkan untuk memudahkan analisis. Tabel ini membantu dalam membandingkan performa model untuk berbagai nilai k, sehingga kita dapat memilih nilai k yang memberikan akurasi terbaik dan membuat keputusan yang didasarkan pada data evaluasi yang jelas.

3.6 Analisis Hasil

Pada tahap Analisis Hasil, langkah pertama adalah Evaluasi Hasil, di mana kita menganalisis akurasi model untuk menentukan nilai k terbaik yang memberikan performa tertinggi. Setelah model

K-Nearest Neighbors (K-NN) diuji dengan berbagai nilai k, hasil akurasi yang diperoleh disimpan dalam tabel hasil. Tabel ini memungkinkan kita untuk membandingkan akurasi model secara langsung untuk setiap nilai k yang diuji. Dengan menganalisis tabel hasil tersebut, kita dapat mengidentifikasi nilai k yang menghasilkan akurasi tertinggi, menunjukkan bahwa model bekerja paling baik dengan parameter tersebut. Evaluasi ini penting untuk memastikan bahwa model K-NN yang dipilih adalah yang paling optimal, memberikan prediksi yang paling akurat berdasarkan data yang tersedia. Dengan memanfaatkan tabel hasil, kita dapat membuat keputusan yang lebih informasi mengenai parameter k yang harus digunakan untuk meningkatkan kinerja model dalam aplikasi praktis.

4. PEMBAHASAN

Dalam penelitian ini, dilakukan optimasi nilai K pada algoritma k-Nearest Neighbors (k-NN) untuk menganalisis sentimen ulasan aplikasi Gopay dengan menggunakan Python. Proses awal melibatkan pengambilan data dari Google Play Store melalui teknik data scrapping, di mana ulasan pengguna diunduh dan disiapkan seperti pemberian label untuk rating lebih dari 3 diberi sentiment 1 yang berarti positif dan kurang dari 3 diberi sentiment 0 yang berarti negatif untuk analisis. Data ulasan aplikasi gopay dapat dilihat untuk 5 baris pertama pada Tabel 1.

Tabel 1. Data Ulasan Aplikasi Gopay

username	*rating*	*time*	*content*	*sentiment*
Pengguna Google	5	7/28/2024 11:49	inovatif dan membantu	1
Pengguna Google	5	7/28/2024 11:49	sangat memuaskan	1
Pengguna Google	5	7/28/2024 11:47	juos gampang sekali	1
Pengguna Google	1	7/28/2024 11:46	hati data pribadi anda bisa di sebar ke orang yg tidak bertanggung	0
Pengguna Google	5	7/28/2024 11:41	mantul lgs proses dan betul	1

Setelah proses pra-pemrosesan, teks ulasan yang asli mengalami penyederhanaan untuk meningkatkan akurasi analisis sentimen. Misalnya, dari baris ulasan awal seperti "inovatif dan membantu" dan "sangat memuaskan," hasil pra-pemrosesan menyederhanakannya menjadi "inovatif membantu" dan "memuaskan." Penghilangan kata-kata yang tidak signifikan dan penyederhanaan frasa bertujuan untuk menekankan kata-kata kunci yang lebih relevan dan mengurangi noise dalam data. Proses ini juga mencakup penghilangan kata-kata tidak penting dan penggantian frasa yang lebih panjang dengan istilah yang lebih ringkas, seperti mengubah "hati data pribadi anda bisa di sebar ke orang lain" menjadi "hati data pribadi sebar orang." Tujuannya adalah untuk menghasilkan representasi teks yang lebih bersih dan terfokus, yang pada gilirannya akan meningkatkan efektivitas analisis sentimen yang dilakukan dengan algoritma k-NN. Dengan optimasi nilai K yang tepat, diharapkan model dapat memberikan hasil sentimen yang lebih akurat dan relevan berdasarkan ulasan pengguna aplikasi Gopay. Hasil sebelum dan sesudah pre-procesiing untuk 5 baris pertama dapat dilihat pada Tabel 2 dan Tabel 3.

Tabel 2. Data Sebelum Pre-Processing

content
inovatif dan membantu
sangat memuaskan
juos gampang sekali
hati data pribadi anda bisa di sebar ke orang yg tidak bertanggung
mantul lgs proses dan betul

Tabel 3. Data Sesudah Pre-Processing

p-ISSN: 2460-173X
e-ISSN: 2598-5841

cleaned content
inovatif membantu
memuaskan
juos gampang
hati data pribadi sebar orang yg
bertanggung
mantul lgs proses

Hasil vektorisasi menghasilkan matriks fitur dengan dimensi (67568, 23413), menunjukkan bahwa terdapat 67.568 ulasan yang dikonversi menjadi vektor dengan 23.413 fitur unik. Matriks ini berbentuk sparse, yang artinya sebagian besar nilai dalam matriks adalah nol, mencerminkan bahwa setiap ulasan hanya berisi sejumlah kecil kata dari keseluruhan kosakata yang digunakan. Contoh data matriks fitur menunjukkan baris-baris yang sebagian besar elemen-nya bernilai nol, dengan hanya beberapa nilai non-nol yang menandakan kemunculan kata tertentu dalam ulasan. Kosakata yang digunakan mencakup istilah beragam seperti 'aa', 'aaa', dan 'aaaaaaa', serta kata-kata lokal seperti 'untuk', 'up', dan 'yang', menandakan keberagaman bahasa dalam ulasan yang dianalisis. Target variabel, yaitu label sentimen, menunjukkan nilai-nilai seperti 1 dan 0 yang merepresentasikan sentimen positif dan negatif, berturut-turut. Beberapa contoh nilai target adalah 1, 1, 1, 0, dan 1, menunjukkan bahwa sebagian besar ulasan dalam dataset ini memiliki sentimen positif. Vektorisasi ini sangat penting untuk menyediakan representasi numerik yang memungkinkan algoritma k-NN menghitung jarak antara ulasan dan melakukan klasifikasi sentimen dengan lebih akurat. Proses ini adalah dasar bagi optimasi nilai K dalam k-NN, yang bertujuan untuk meningkatkan keandalan dan ketepatan dalam menganalisis sentimen dari ulasan aplikasi. Untuk hasil bentuk matriks fitur, contoh data matrix fitur untuk 5 baris pertama, contoh kosakata yang digunakan, dan nilai target sentimen untuk 5 baris pertama dapat dilihat pada Gambar 2, Gambar 3 dan Gambar 4.

```
Bentuk matriks fitur: (67568, 23413)
Contoh data matriks fitur:
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

Gambar 2. Bentuk matriks fitur dan contoh data matriks fitur

```
Contoh Kosakata yang digunakan:
['aa' 'aaa' 'aaaaaaa' ... 'untuk' 'up' 'yang']
```

Gambar 3. Contoh kosakata yang digunakan

```
Nilai dari target sentiment:
0    1
1    1
2    1
3    0
4    1
```

Gambar 4. Nilai dari target sentiment

Hasil analisis menunjukkan bahwa akurasi model k-NN cenderung menurun saat nilai k meningkat. Pada nilai k=3, model mencapai akurasi tertinggi sebesar 0.898. Namun, akurasi ini menurun secara bertahap seiring dengan bertambahnya nilai k, mencapai 0.867 pada k=49. Penurunan akurasi ini dapat diartikan bahwa model k-NN dengan nilai k yang lebih besar mungkin terlalu umum dalam menilai data, sehingga mengurangi kemampuannya untuk menangkap pola-

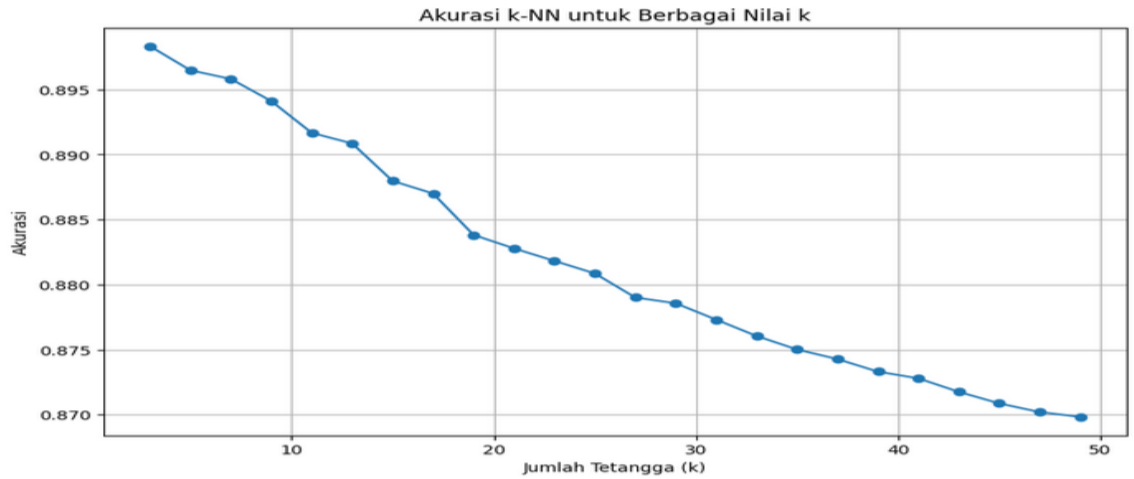
p-ISSN: 2460-173X
e-ISSN: 2598-5841

pola spesifik yang ada dalam ulasan. Dengan demikian, nilai k yang lebih kecil terbukti lebih efektif dalam konteks ini, memberikan indikasi bahwa memilih nilai k yang lebih kecil dapat menghasilkan performa yang lebih baik dalam analisis sentimen ulasan aplikasi. Temuan ini menekankan pentingnya melakukan pemilihan parameter yang cermat untuk mengoptimalkan hasil dari algoritma k-NN dalam aplikasi praktis. Hasil analisis dapat dilihat pada Tabel 4.

Tabel 4. Hasil Analisis

Nilai k	*Akurasi*	*Nilai k*	*Akurasi*
3	0.898	27	0.879
5	0.897	29	0.879
7	0.896	31	0.877
9	0.894	33	0.876
11	0.892	35	0.875
13	0.891	37	0.874
15	0.888	39	0.873
17	0.887	41	0.873
19	0.884	43	0.872
21	0.883	45	0.871
23	0.882	47	0.870
25	0.881	49	0.870

Visualisasi yang ditampilkan menggambarkan hubungan antara jumlah tetangga (k) dan akurasi model k-NN. Data menunjukkan bahwa akurasi model cenderung menurun seiring dengan meningkatnya nilai k. Misalnya, akurasi tertinggi mencapai 0.898 pada k=3, namun mulai menurun secara bertahap hingga mencapai 0.870 pada k=47 dan k=49. Grafik ini, yang memplot akurasi terhadap nilai k dengan marker berbentuk lingkaran, mengindikasikan bahwa nilai k yang lebih kecil memberikan hasil yang lebih baik dalam hal akurasi. Penurunan akurasi dengan meningkatnya k mungkin disebabkan oleh fakta bahwa model k-NN dengan nilai k yang lebih tinggi cenderung lebih halus dan kurang responsif terhadap variasi data, sehingga dapat mengurangi kemampuan model dalam menangkap pola spesifik dalam data ulasan. Oleh karena itu, hasil ini menunjukkan pentingnya memilih nilai k yang optimal untuk mencapai akurasi terbaik dalam analisis sentimen. Hasil analisis di visualisasikan seperti terlihat pada Gambar 5.



Gambar 5. Grafik Hasil Analisis

5. KESIMPULAN

Penelitian ini memberikan wawasan yang berharga tentang bagaimana nilai k mempengaruhi performa algoritma k-NN dalam analisis sentimen ulasan aplikasi Gopay dengan data yang

diperoleh melalui teknik scraping dari Google Play Store. Dengan mengidentifikasi nilai k yang optimal, penelitian ini dapat membantu pengembang aplikasi dan perusahaan dalam menerapkan algoritma k-NN dengan lebih efektif untuk menganalisis umpan balik pengguna. Selain itu, penelitian ini juga menyarankan bahwa pemilihan nilai k yang lebih kecil dapat memberikan hasil yang lebih baik dibandingkan dengan nilai k yang lebih besar, terutama dalam konteks analisis sentimen aplikasi. Kesimpulan ini dapat diterapkan pada studi serupa di masa depan untuk meningkatkan pemahaman dan analisis sentimen dalam berbagai domain aplikasi.

Dengan pemahaman yang lebih baik tentang cara kerja algoritma k-NN dan dampak dari nilai k terhadap akurasi, penelitian ini membuka peluang untuk eksplorasi lebih lanjut dalam pengembangan teknik analisis sentimen yang lebih efektif. Penelitian ini juga menunjukkan pentingnya pemilihan parameter dalam algoritma pembelajaran mesin dan bagaimana optimasi dapat meningkatkan kinerja model. Ke depannya, teknik ini dapat diterapkan pada dataset yang lebih besar dan lebih beragam, serta diperluas untuk mencakup metode optimasi lainnya untuk meningkatkan hasil analisis sentimen.

DAFTAR PUSTAKA

[1] S. W. Iriananda, R. W. Budiawan, and A. Y. Rahman, "Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan SVM, Grid Search Dan Kombinasi N-Gram Optimizing Sentiment Classification Of Mobile Game User Reviews Using SVM, Grid Search And N-Gram Combinations," vol. 11, no. 4, 2024, doi: 10.25126/jtiik.1148244.

[2] M. A. Yamin, K. Kusnadi, and L. Bayuaji, "Optimasi Algoritma Support Vector Machine (SVM) Dengan Menggunakan Feature Selection Gain Ratio Untuk Analisis Sentimen," *INOVTEK Polbeng - Seri Inform.*, vol. 9, no. 1, pp. 326–340, 2024, doi: 10.35314/isi.v9i1.4197.

[3] D. R. Fernandes, N. J. P. Hasan, and N. Wijaya, "Optimasi Akurasi Sentimen Komentar Xiaomi SU7 di YouTube Menggunakan Naive Bayes dan Chi-Square," *J. Softw. Eng. Comput. Intell.*, vol. 2, no. 01, pp. 17–25, 2024, doi: 10.36982/jseci.v2i01.4099.

[4] V. Arinal and B. S. Purnomo, "Optimasi Metode Decision Tree Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Review Game GTA V Roleplay," *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 457–461, 2023, [Online]. Available: https://doi.org/10.55338/saintek.v5i1.1371

[5] I. P. D. W. Darmawan, G. A. Pradnyana, and I. B. N. Pascima, "Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram," *SINTECH (Science Inf. Technol. J.*, vol. 6, no. 1, pp. 58–67, 2023, doi: 10.31598/sintechjournal.v6i1.1245.

[6] P. Arsi, R. Wahyudi, and R. Waluyo, "Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 231–237, 2021, doi: 10.29207/resti.v5i2.2698.

Biodata Penulis

***Anas Nasrulloh, M.Kom.*,** merupakan lulusan S2 Universitas Budi Luhur Jakarta jurusan Rekayasa Komputer Terapan. Beliau adalah dosen Sistem Informasi di Institut Teknologi Tangerang Selatan (ITTS). Fokus penelitiannya meliputi data science, machine learning, dan analisis sentimen menggunakan Python. Jurnal ini ditulis sebagai kontribusi dalam pengembangan penelitian kecerdasan buatan dan analisis data teks di Indonesia. Sinta ID: 6759995; Email: anas@itts.ac.id

Tubagus Toifur, M.Kom., seorang akademisi dan praktisi di bidang Teknologi Informasi dengan latar belakang pendidikan Ilmu Komputer, ia saat ini menjabat sebagai Wakil Rektor Bidang II di Institut Teknologi Tangerang Selatan (ITTS), sekaligus aktif mengajar mata kuliah terkait multimedia, pengolahan citra digital, etika profesi, dan keamanan siber. Sebelum berkiprah di dunia pendidikan tinggi, ia memiliki pengalaman profesional yang luas di berbagai instansi pemerintah dan BUMD, termasuk sebagai IT Support di Dinas Pertambangan dan Energi Provinsi Banten serta Kepala Divisi IT di PT PITS Tangerang Selatan. Berkontribusi dalam beberapa publikasi ilmiah dan penulisan buku, di antaranya sebagai penulis pada Buku Pegangan Penanganan Insiden Siber (2023) dan Panduan Migrasi Jaringan IPv4 ke IPv6 (2024), serta berbagai artikel teknis yang dipublikasikan di laman instansi pemerintah. Email: tubagus@itts.ac.id

Aolia Ikhwanuddin, M.Kom., adalah akademisi dan praktisi teknologi informasi yang berfokus pada jaringan komputer, cybersecurity, serta pengembangan Artificial Intelligence (AI). Ia aktif sebagai dosen sekaligus Dekan Fakultas Ilmu Komputer di Institut Teknologi Tangerang Selatan (ITTS), serta berperan sebagai staf ahli di Dinas Komunikasi dan Informatika Kota Tangerang Selatan. Dengan pengalaman di dunia akademik dan pemerintahan, ia aktif mendorong transformasi digital, penguatan kapasitas teknologi, dan tata kelola layanan publik berbasis digital. Selain mengajar, ia juga aktif menulis berbagai buku dan referensi di bidang teknologi informasi, khususnya jaringan komputer, keamanan siber,Internet Of Things dan analisis data. Melalui buku ini, penulis berharap dapat memberikan wawasan yang bermanfaat bagi pelajar, mahasiswa, pendidik, dan praktisi agar lebih mudah memahami perkembangan teknologi secara praktis, sistematis, dan relevan dengan kebutuhan era digital.

Muhamad Yusuf, M.Kom., adalah seorang akademisi, praktisi teknologi informasi, dan pengembang sistem aplikasi pemerintahan yang telah berkecimpung dalam dunia IT selama lebih dari satu dekade. Saat ini, ia menjabat sebagai Kepala Program Studi Teknologi Informasi di Institut Teknologi Tangerang Selatan, serta aktif mengajar sebagai dosen pada fakultas teknik di kampus yang sama. Selain di dunia akademik, Yusuf juga berperan sebagai Tenaga Ahli di Dinas Komunikasi dan Informatika Kota Tangerang Selatan, di mana ia terlibat dalam pengembangan berbagai sistem informasi yang mendukung pelayanan publik dan transformasi digital pemerintah daerah. Email: yusuf@itts.ac.id

***Ibnu Mas'ud, M.Kom.*,** Seorang akademisi dan praktisi di bidang Ilmu Komputer yang memiliki dedikasi tinggi terhadap pengembangan teknologi informasi. Dengan latar belakang pendidikan Magister Komputer, beliau memiliki spesialisasi dalam analisis sistem, pengelolaan data, serta implementasi teknologi terkini untuk mendukung solusi bisnis dan pendidikan. Saat ini, beliau aktif mengajar sebagai dosen di Universitas Islam Negeri Sultan Maulana Hasanuddin Banten dan turut berkontribusi sebagai pengajar di beberapa perguruan tinggi lainnya, termasuk Institut Teknologi Tangerang Selatan. Email: ibnumasud@uinbanten.ac.id, ibnu@itts.ac.id