

ANALISIS SENTIMEN APLIKASI PADI UMKM DENGAN PENINGKATAN KINERJA ALGORITMA KNN

Mesi Febima¹, Unang Solihin², Lena Magdalena³, Muhammad Hatta⁴, Marsani Asfi⁵, Stefanny Christina⁶

Universitas Catur Insan Cendekia

Jl. Kesambi 202, Kota Cirebon, Jawa Barat Tlp : (0231) 220250

e-mail: mesi.febima@cic.ac.id¹, unang.solihin@cic.ac.id², lena.magdalena@cic.ac.id³,
muhammad.hatta@cic.ac.id⁴, marsani.asfi@cic.ac.id⁵, stefanny.christina@cic.ac.id⁶

Abstrak

Dalam era digital, Usaha Mikro, Kecil, dan Menengah (UMKM) dituntut untuk memanfaatkan teknologi guna memperluas pasar dan meningkatkan daya saing. Salah satu inovasi yang mendukung hal tersebut adalah aplikasi PaDi UMKM, platform hasil inisiatif Kementerian BUMN yang mempertemukan BUMN dengan produk-produk berkualitas dari UMKM di seluruh Indonesia. Keberhasilan aplikasi ini tidak hanya ditentukan oleh fungsionalitasnya, tetapi juga oleh persepsi dan pengalaman pengguna yang tercermin melalui ulasan di Google Playstore. Untuk memahami persepsi tersebut, dilakukan analisis sentimen menggunakan pendekatan Natural Language Processing (NLP) dan pembelajaran mesin. Penelitian ini bertujuan untuk menganalisis sentimen positif dan negatif terhadap aplikasi PaDi UMKM dengan membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Improved K-Nearest Neighbor (IKNN). Proses penelitian meliputi pengumpulan data ulasan, pra-proses teks, pembagian data latih dan uji, penerapan algoritma, serta evaluasi hasil klasifikasi menggunakan confusion matrix. Berdasarkan hasil confusion matrix, nilai $K = 5$ memberikan performa terbaik dibandingkan $K = 3$, $K = 7$, dan $K = 9$, dilihat dari peningkatan nilai Precision 50 % positif dan 67% negatif, Recall 80% positif dan 33% negatif, F1-Score 62% positif dan 44% negatif, dan Accuracy 10% positif dan 12% negatif. Sebaliknya, algoritma IKNN menunjukkan peningkatan kinerja yang signifikan dengan nilai Precision, Recall, F1-Score, dan Accuracy mencapai 100% pada seluruh variasi K . Hal ini membuktikan bahwa peningkatan metode KNN melalui pendekatan IKNN mampu menghasilkan klasifikasi sentimen yang jauh lebih akurat dan konsisten. Dengan demikian, IKNN terbukti lebih efektif dan dapat menjadi acuan dalam pengembangan sistem analisis sentimen berbasis kecerdasan buatan di masa mendatang

Kata kunci: PaDi, UMKM, Analisis Sentimen, KNN, IKNN

Abstract

In the digital era, Micro, Small, and Medium Enterprises (MSMEs) are required to utilize technology to expand their markets and enhance competitiveness. One innovation that supports this effort is the PaDi UMKM application, a platform initiated by the Ministry of State-Owned Enterprises (BUMN) that connects state-owned companies with high-quality MSME products across Indonesia. The success of this application is not only determined by its functionality but also by user perceptions and experiences reflected in reviews on the Google Play Store. To understand these perceptions, a sentiment analysis was conducted using a Natural Language Processing (NLP) and machine learning approach. This study aims to analyze positive and negative sentiments toward the PaDi UMKM application by comparing the performance of the K-Nearest Neighbor (KNN) and Improved K-Nearest Neighbor (IKNN) algorithms. The research process includes collecting user review data, preprocessing text, splitting the dataset into training and testing sets, applying the algorithms, and evaluating classification results using a confusion matrix. Based on the confusion matrix results, $K = 5$ provided the best performance compared to $K = 3$, $K = 7$, and $K = 9$, with improvements in Precision (50% for positive and 67% for negative), Recall (80% for positive and 33% for negative), F1-Score (62% for positive and 44% for negative), and Accuracy (10% for positive and 12% for negative). In contrast, the IKNN algorithm demonstrated a significant improvement, achieving 100% in Precision, Recall, F1-Score, and Accuracy across all K values. This indicates that enhancing the KNN method through the IKNN approach produces far more accurate and consistent sentiment classifications. Therefore, IKNN is proven to be more effective and can serve as a reference for future development of artificial intelligence-based sentiment analysis systems

Keywords: PaDi, MSMEs, Sentiment Analysis, KNN, IKNN

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong berbagai sektor untuk beradaptasi, termasuk sektor Usaha Mikro, Kecil, dan Menengah (UMKM). Dalam perkembangannya UMKM tidak hanya menjadi tulang punggung domestik, tetapi juga memiliki potensi besar untuk memasuki pasar Internasional [1]. Persaingan usaha yang semakin kompetitif menuntut para pelaku UMKM untuk terus berinovasi agar dapat mempertahankan eksistensi dan memperluas jangkauan bisnisnya [2]. Dalam era digital, UMKM dituntut untuk mampu memanfaatkan platform daring guna memperluas pasar, meningkatkan layanan, serta memperkuat hubungan dengan pelanggan. Salah satu bentuk inovasi yang mendukung hal tersebut adalah aplikasi PaDi UMKM. Aplikasi PaDi UMKM adalah platform yang dirancang untuk mempertemukan BUMN dengan produk-produk berkualitas dari UMKM di seluruh Indonesia.

Platform ini merupakan inisiatif dari Kementerian BUMN yang bekerja sama dengan sejumlah BUMN, dengan tujuan untuk menggerakkan ekosistem UMKM secara digital, sistematis, dan berkelanjutan dalam skala nasional. Melalui platform ini, UMKM dapat memperoleh akses yang lebih luas terhadap pasar dan peluang kerja sama dengan perusahaan besar, sehingga mendorong pertumbuhan ekonomi dan digitalisasi sektor UMKM di Indonesia. Namun demikian, keberhasilan suatu aplikasi tidak hanya ditentukan oleh fungsionalitasnya, tetapi juga oleh respon dan persepsi pengguna. Ulasan dan komentar pengguna di media sosial atau platform aplikasi menjadi sumber informasi penting yang mencerminkan tingkat kepuasan, keluhan, maupun saran perbaikan. Untuk memahami hal tersebut secara efektif, diperlukan pendekatan analisis sentimen, yaitu teknik pengolahan teks untuk mengidentifikasi opini positif, negatif, atau netral dari pengguna terhadap suatu produk atau layanan.

Ulasan positif mengandung informasi komentar atau opini pengguna yang menunjukkan kepuasan, pengalaman baik, atau pandangan yang mendukung terhadap suatu produk, layanan, atau aplikasi PaDi UMKM. Ulasan negatif merupakan informasi yang dapat menunjukkan masalah teknis, performa aplikasi yang kurang optimal, bug, kesulitan dalam proses transaksi, atau respon sistem yang lambat, yang semuanya dapat menjadi masukan penting bagi pengembang aplikasi. *Google Play Store* merupakan platform utama di mana pengguna dapat memberikan ulasan positif dan negatif mengenai aplikasi PaDi UMKM. Banyaknya ulasan yang masuk setiap hari, analisis manual menjadi tidak praktis dan memakan waktu. Oleh karena itu, untuk mengekstraksi informasi berguna dari sejumlah besar ulasan positif dan negatif diperlukan teknik analisis yang efisien dan akurat.

Penelitian ini menggunakan *Natural Language Processing* (NLP) dan *machine learning* untuk metode efektif untuk menganalisis alasan pengguna. *Natural Language Processing* (NLP) adalah cabang ilmu komputer yang berkaitan dengan pemahaman dan pemrosesan bahasa manusia secara alami oleh komputer [3]. NLP memungkinkan komputer memahami, menerjemahkan, dan mengelola bahasa yang digunakan manusia, sedangkan *machine learning* memungkinkan sistem belajar dari informasi yang diberikan dan terus meningkatkan kemampuannya seiring berjalannya waktu.

Penelitian ini menggunakan *machine learning* dalam NLP bertujuan untuk meningkatkan akurasi dan efisiensi analisis sentimen dari ulasan positif dan negatif aplikasi PaDi UMKM. Proses NLP yang digunakan untuk memproses dan menganalisis data mencakup beberapa tahapan, yaitu mengumpulkan ulasan dari pengguna, mempersiapkan data, menerapkan algoritma pembelajaran mesin, serta mengevaluasi hasil dari model yang telah dibuat. Analisis sentimen adalah proses yang memanfaatkan NLP untuk secara otomatis mengetahui sikap, pendapat, pandangan, serta perasaan seseorang dari berbagai jenis data seperti teks, suara, tweet, dan sumber informasi lainnya [4]. Analisis sentimen dalam ulasan Aplikasi PaDi UMKM tidak hanya memberikan wawasan mengenai tingkat kepuasan dan ketidakpuasan pengguna, tetapi juga menawarkan sudut pandang yang lebih mendalam untuk memahami preferensi pengguna serta keefektifan penggunaan PaDi UMKM sebagai platform digital yang mempertemukan pelaku UMKM dengan BUMN dalam mendukung pertumbuhan ekonomi nasional secara berkelanjutan.

Klasifikasi adalah bentuk dasar dari analisa data dan sering juga dikenal sebagai teknik untuk menentukan kategori kelas dari data-data yang sudah ada [5]. Klasifikasi merupakan bagian dari algoritma *supervised learning*. Model klasifikasi dalam analisis sentimen merupakan salah satu algoritma yang efektif untuk mengkategorikan ulasan pengguna ke dalam pengguna ke dalam klasifikasi ulasan positif dan ulasan negatif. Dalam penelitian ini mengimplementasikan algoritma K-Nearest Neighbor (KNN) digunakan sebagai metode klasifikasi sentimen. Algoritma K-Nearest Neighbor atau KNN merupakan algoritma dalam melakukan klasifikasi objek berdasarkan jarak data pembelajaran baru berdasarkan K tetangga yang paling dekat [6]. Algoritma KNN dikenal karena kesederhanaan dan efektivitasnya dalam mengelompokkan data berdasarkan kedekatan antar titik dalam ruang fitur.

Meskipun demikian, performa KNN sangat dipengaruhi oleh pemilihan parameter, seperti nilai k dan metode pembobotan jarak. Oleh karena itu, dalam penelitian ini dilakukan peningkatan kinerja

algoritma KNN dengan menerapkan metode *Improved K-Nearest Neighbor* (IKNN) untuk menganalisis sentimen terhadap aplikasi PaDi UMKM berdasarkan komentar pengguna yang diperoleh dari platform *Google Playstore*. Pendekatan ini diharapkan mampu meningkatkan akurasi klasifikasi sentimen dengan mempertimbangkan bobot jarak yang lebih adaptif serta distribusi data yang lebih optimal. Selain itu, tujuan dari penelitian ini adalah melakukan perbandingan antara algoritma KNN dan IKNN dalam menganalisis sentimen pengguna terhadap aplikasi PaDi UMKM, guna mengetahui algoritma mana yang memiliki performa lebih baik dalam hal akurasi, presisi, dan keandalan hasil klasifikasi.

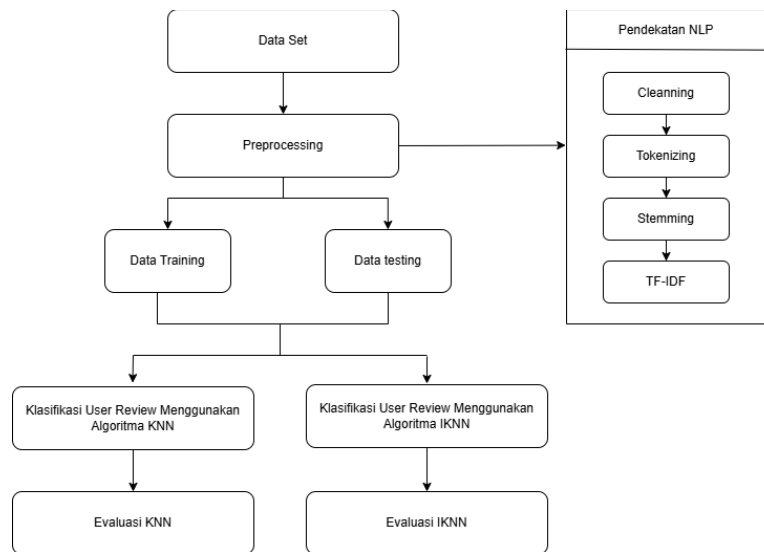
Penelitian terkait analisis sentimen ulasan aplikasi telah dilakukan sebelumnya. Beberapa penelitian terdahulu yang berkaitan dengan analisis sentimen adalah penelitian yang dilakukan oleh [7] tentang analisis sentimen tentang aplikasi SAMPEAN Kota Cirebon menggunakan NLP dan KNN, berdasarkan hasil penelitian yang dilakukan dari 16 data uji, diperoleh hasil bahwa terdapat 10 data yang diklasifikasikan sebagai ulasan teknis dan 6 data yang diklasifikasikan sebagai ulasan pengalaman pengguna. Hasil ini menunjukkan bahwa klasifikasi sentimen pengguna terhadap aplikasi SAMPEAN Kota Cirebon lebih banyak pada ulasan teknis dibandingkan dengan ulasan pengalaman pengguna. Selanjutnya penelitian yang dilakukan oleh [8] tentang klasifikasi sentimen ojek online di Indnesia, berdasarkan hasil penelitian menunjukkan hasil percobaan model klasifikasi dengan algoritma KNN dan IKNN menunjukkan akurasi pada transportasi online Grab adalah 77,77% sedangkan untuk aplikasi Gojek sebesar 66,27% pada algoritma KNN. Sedangkan hasil algoritma IKNN Grab memiliki akurasi lebih tinggi yaitu 77,39% dan Gojek memiliki nilai akurasi sebesar 48,92%.

Penelitian ini bertujuan untuk menghasilkan model analisis sentimen yang mampu memberikan gambaran persepsi masyarakat terhadap aplikasi PaDi UMKM dengan membandingkan algoritma KNN dengan IKNN, sekaligus menjadi dasar bagi pengembang dan pengelola untuk melakukan peningkatan kualitas layanan. Dengan demikian, hasil penelitian ini tidak hanya berkontribusi dalam bidang data mining dan kecerdasan buatan, tetapi juga memberikan manfaat nyata bagi penguatan ekosistem digital UMKM di Indonesia.

2. METODE PENELITIAN

Dalam penelitian ini, proses penelitian dibagi menjadi empat tahap utama. Tahap pertama adalah pengumpulan data ulasan pengguna aplikasi PaDi UMKM yang diperoleh dari *platform Google Play Store*. Data ulasan ini berisi opini dan tanggapan pengguna yang mencerminkan kepuasan maupun ketidakpuasan terhadap layanan aplikasi. Tahap kedua adalah pengolahan data menggunakan pendekatan *Natural Language Processing* (NLP) untuk melakukan proses text preprocessing seperti *cleansing*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, sehingga data teks menjadi lebih terstruktur dan siap digunakan untuk analisis sentimen.

Tahap ketiga adalah proses klasifikasi sentimen, di mana data dibagi menjadi dua bagian, yaitu data latih dan data uji. Kedua dataset ini digunakan untuk mengklasifikasikan ulasan pengguna menjadi dua kategori utama, yaitu sentimen positif dan sentimen negatif terhadap aplikasi PaDi UMKM. Proses klasifikasi dilakukan dengan menggunakan dua algoritma berbeda, yaitu K-Nearest Neighbor (KNN) dan Improved K-Nearest Neighbor (IKNN), guna melihat perbedaan performa dan akurasi di antara keduanya. Tahap terakhir adalah evaluasi hasil klasifikasi, yang bertujuan untuk mengukur tingkat *akurasi*, *presisi*, dan *recall* dari masing-masing algoritma. Hasil evaluasi ini digunakan untuk menentukan algoritma mana yang memiliki kinerja terbaik dalam menganalisis sentimen pengguna terhadap aplikasi PaDi UMKM. Tahapan penelitian secara keseluruhan dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

1. **Data Set**
Penelitian ini menggunakan teknik crawling untuk mengumpulkan data set user reviews pada aplikasi PaDi UMKM yang diperoleh pada link ‘com.telkom.padiumkm’. Play Store. Data Set yang diperoleh sejumlah 109 data *user reviews*.
2. **Preprocessing**
Preprocessing data adalah langkah penting dalam penelitian yang bertujuan untuk membersihkan, merapikan, dan mengubah data mentah menjadi format yang lebih sesuai untuk dianalisis lebih lanjut. Tujuan dari *preprocessing* data adalah untuk membersihkan data teks dengan cara yang tepat agar bisa dianalisis di masa depan. Penelitian ini menggunakan tahapan pra-pemrosesan data dengan pendekatan Natural Language Processing (NLP).
3. **Natural Language Processing (NLP)**
Natural Language Processing (NLP) adalah cabang ilmu komputer yang berkaitan dengan pemahaman dan pemrosesan bahasa manusia secara alami oleh komputer [3]. NLP memungkinkan komputer memahami, menerjemahkan, dan mengelola bahasa yang digunakan manusia. Tahapan-tahapan dari NLP terdiri dari:
 - a. **Cleanning**
Proses tahapan *cleanning* dilakukan dengan menghapus karakter yang tidak penting seperti emoji, tanda baca, *URL*, *hashtag*, dan angka dari item yang tidak relevan sehingga membuat data lebih bersih dan mudah dilakukan proses pada langkah selanjutnya. Tujuan dari *cleanning* ini yakni untuk membersihkan data dari *noise*, informasi yang tidak relevan, dan kesalahan sehingga data yang digunakan dalam analisis lebih bersih, akurat, dan dapat diandalkan dalam analisis.
 - b. **Tokenizing**
Tokenizing merupakan langkah kedua dalam pendekatan NLP di mana tahap ini akan memecah teks menjadi token, seperti kata-kata atau kalimat. Ini memudahkan untuk melanjutkan pemrosesan. Library yang digunakan dalam proses ini adalah NLTK (*Natural Language Toolkit*).
 - c. **Stemming**
Proses *Stemming* berfungsi untuk mengkonversi kata-kata menjadi bentuk kata dasar. Contohnya kata “berlari-lari” dan “berlari” dapat disederhanakan menjadi “lari”. *Stemming* berperan dalam mengurangi perbedaan kata yang memiliki makna yang sama. Library yang digunakan dalam proses stemming adalah sastrawi yang mampu mengubah kata-kata terimbu dalam Bahasa Indonesia ke bentuk kata dasar.
 - d. **TF-IDF**
TF-IDF merupakan proses yang digunakan untuk melakukan transformasi data dari sebuah teks menjadi sebuah numerik guna dilakukan untuk pembobotan pada tiap kata [9]. Teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) memperkirakan seberapa penting sebuah kata dalam dokumen atau koleksi dokumen. Rumus kalkulasi TF-IDF

menggabungkan frekuensi kata dalam dokumen dengan inversi frekuensi kata dalam keseluruhan dokumen, sehingga memberikan bobot yang lebih tinggi pada kata-kata yang jarang muncul tapi memiliki makna penting [4].

4. Klasifikasi *User Review* Menggunakan Algoritma KNN dan IKNN

Algoritma K-Nearest Neighbor atau KNN merupakan algoritma dalam melakukan klasifikasi objek berdasarkan jarak data pembelajaran baru berdasarkan K tetangga yang paling dekat [10]. Algoritma KNN dikenal karena kesederhanaan dan efektivitasnya dalam mengelompokkan data berdasarkan kedekatan antar titik dalam ruang fitur. Klasifikasi dalam algoritma K-Nearest Neighbor dilakukan beberapa tahapan menurut [11], di antaranya:

1. Tentukan jumlah k
2. Hitung jarak antara titik data yang diinginkan dan k tetangga menggunakan jarak euclidean yang dapat dihitung dengan persamaan (1)

$$D(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \quad (1)$$

Di mana p dan q merupakan titik data

3. Pilih dan mengurutkan kelompok-kelompok k tetangga terdekat yang memiliki jarak euclidean yang terkecil
4. Hitung jumlah titik data dari setiap kelas k tetangga yang dipilih pada tahapan sebelumnya
5. Posisikan titik data baru ke dalam kategori di mana jumlah tetangga maksimum.

5. Improved IKNN

Perbedaan antara Improved KNN dan metode KNN terletak pada nilai k yang ditentukan secara berbeda. Untuk mencapai tingkat akurasi yang tinggi dalam proses klasifikasi dokumen, nilai k pada KNN harus ditentukan secara tepat dan presisi [8]. Penentuan nilai k tersebut kemudian dihitung untuk mendapatkan nilai k(n) yang baru. Penentuan nilai k(n) yang baru dapat dihitung menggunakan persamaan (2).

$$n = \frac{k * N(cm)}{\text{Maks}[N(cm)]_{j=1..Nc}} \quad (2)$$

Keterangan:

n = Nilai k baru

k = Nilai k yang ditetapkan

Maks[N(cm)]_{j=1..Nc} = Banyak dokumen latih terbanyak pada semua kategori

Selanjutnya dilakukan perhitungan peluang dari dokumen uji X termasuk dengan dokumen latih dj sebanyak nilai n tetangga untuk setiap kategori pada dokumen X pada dokumen latih dj sebanyak nilai n tetangga untuk trainingset. Persamaan berikut dapat digunakan untuk menghitung peluang dari dokumen uji X pada kategori m.

$$\frac{\sum_{dj \in \text{top_n_kNN}(cm)} \text{sim}(x, dj) y(dj, cm)}{\sum_{dj \in \text{top_n_kNN}(cm)} \text{sim}(x, dj)} \quad (3)$$

Keterangan:

P(x, cm) = Probabilitas dokumen X

Sim(x, dj) = Kemiripan antara dokumen X dengan dokumen latih

Top_n_kNN = Nilai n terbaik tetangga

Y(dj, cm) = Fungsi atribut yang memenuhi dari salah satu kategori

6. Akurasi

Untuk mengevaluasi performa model klasifikasi, terdapat confusion matrix yang membandingkan hasil klasifikasi aktual dengan prediksi yang diberikan oleh model. Ada empat nilai penting dalam matrix confusion: Nilai Positif Benar (TP), Nilai Negatif Benar (TN), Nilai Negatif Benar (FP),

dan Nilai Negatif Benar (FN). Proses ini melibatkan perhitungan nilai accuracy, precision, recall, dan f1-score. Nilai precision digunakan untuk mengukur seberapa akurat model dalam memprediksi data positif, sedangkan nilai recall digunakan untuk menghitung seberapa baik model dalam menemukan data positif, dan f1-score digunakan untuk mengukur harmonic mean atau keseimbangan antara nilai precision dan recall [12]. Evaluasi kinerja ini digunakan sebagai parameter untuk mengukur seberapa akurat implementasi metode yang digunakan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \quad (7)$$

3. HASIL DAN PEMBAHASAN

3.1. Data Set

Teknik *crawling* yang dilakukan pada proses pengumpulan yang diperoleh pada link ‘com.telkom.padiumkm” pada *Google Playstore* menghasilkan sentimen sebanyak 109 *record* dengan sample sentimen seperti tampak dalam Tabel 1 berikut.

Tabel 1. Data Set Hasil Scrapping

No	Sentimen
1	pengen rasa nya ada ecommerce di miliki oleh negara..namun potongan nya jngn terlalu besar seperti milik swasta.slama ini ecommerce di Indonesia di kuasai swasta dan investor asing..sudah saat nya ecommerce dimiliki oleh negara untuk mendukung Umkm. selama ini sya ikut jualan di platform swasta sperti si ijo si biru si oren dan bahkan tok tok.nmun potongan nya sekarang terlalu besar.penjualan 25 RB.yg kita terima paling 17k setelah dipotong.17k itu kita cuma untung 3k aja..ayo BUMN bikin segera.
2	chat penjual TDK ada tanggapan Krn offline sepanjang HR...akibatnya pesan loading nya g sampe sampe. jelek respon chatnya. sdh dibayar barang blm dikirim Krn respon penjual dr chatnya jelek.sdh dibayar tgl 9 sampe dengan tanggal 14 pun tidak dikirim atau direfund uang nya. ini aplikasi penipuan. hati hati jangan transaksi dengan aplikasi ini
3	Bagus digunakan
4	Server error terus, udah bayar ga kedetect!! pesanannya yg udh saya bayar malah tidak ada di history
5	server gangguan terus, berat berat, padahal di wajibkan beli di sini
...
108	Bermanfaat
109	Hidup UMKM

3.2. Preprocessing Data

Tahapan preprocessing menggunakan pendekatan NLP (*Natural Language Processing*) atau dikenal dengan *text processing*. Adapun tahap yang dilakukan sebagai berikut: *Cleanning*, tokenizing, *stemming*, stopword, dan terakhir TF-IDF. Tahapan tersebut secara sistematis dilakukan agar memperoleh output data yang optimal.

1. *Cleanning*

Proses tahapan *cleanning* dilakukan dengan menghapus karakter yang tidak penting seperti emoji, tanda baca, URL, hashtag, dan angka dari item yang tidak releva sehingga membuat data lebih bersih dan mudah dilakukan proses pada langkah selanjutnya. Tabel 2 berikut merupakan simulasi teks sebelum dan sesudah proses *cleanning* data.

Tabel 2. Proses *Cleanning*

No	Sentimen sebelum <i>cleanning</i>	Sentimen sesudah <i>cleanning</i>
1	pengen rasa nya ada ecommerce di miliki oleh negara..namun potongan nya jngn terlalu besar seperti milik swasta.slama ini ecommerce di Indonesia di kuasai swasta dan investor asing..sudah saat nya ecommerce dimiliki oleh negara untuk mendukung Umkm. selama ini sya ikut jualan di platform swasta sperti si ijo si biru si oren dan bahkan tok tok.nmun potongan nya sekarang terlalu besar.penjualan 25 RB.yg kita terima paling 17k setelah dipotong.17k itu kita cuma untung 3k aja..ayo BUMN bikin segera.	pengen rasa nya ada ecommerce di miliki oleh negara..namun potongan nya jngn terlalu besar seperti milik swasta.slama ini ecommerce di indonesia di kuasai swasta dan investor asing..sudah saat nya ecommerce dimiliki oleh negara untuk mendukung umkm. selama ini sya ikut jualan di platform swasta sperti si ijo si biru si oren dan bahkan tok tok.nmun potongan nya sekarang terlalu besar.penjualan 25 rb.yg kita terima paling 17k setelah dipotong.17k itu kita cuma untung 3k aja..ayo bumN bikin segera.
2	chat penjual TDK ada tanggapan Krn offline sepanjang HR...akibatnya pesan loading nya g sampe sampe. jelek respon chatnya. sdh dibayar barang blm dikirim Krn respon penjual dr chatnya jelek.sdh dibayar tgl 9 sampe dengan tanggal 14 pun tidak dikirim atau direfund uang nya. ini aplikasi penipuan. hati hati jangan transaksi dengan aplikasi ini	chat penjual tdk ada tanggapan krn offline sepanjang hr...akibatnya pesan loading nya g sampe sampe. jelek respon chatnya. sdh dibayar barang blm dikirim krn respon penjual dr chatnya jelek.sdh dibayar tgl 9 sampe dengan tanggal 14 pun tidak dikirim atau direfund uang nya. ini aplikasi penipuan. hati hati jangan transaksi dengan aplikasi ini
3	Bagus digunakan	bagus digunakan
4	Server error terus, udah bayar ga kedetect!! pesannya yg udh saya bayar malah tidak ada di history	server error terus, udah bayar ga kedetect!! pesannya yg udh saya bayar malah tidak ada di history
5	server gangguan terus, berat berat, padahal diwajibkan beli di sini	server gangguan terus, berat berat, padahal diwajibkan beli di sini
..
108	Bermanfaat	bermanfaat
109	Hidup UMKM	hidup umkm

2. *Tokenizing*

Tokenizing merupakan proses pemecahan sekumpulan karakter dalam suatu teks kedalam satuan kata, pada proses ini juga dapat menghilangkan karakter pembatas, menghapus angka, dan menghapus tanda baca atau karakter yang bukan merupakan huruf.

Tabel 3. Proses *Tokenizing*

No	Sentimen sebelum <i>tokenizing</i>	Sentimen sesudah <i>tokenizing</i>
1	pengen rasa nya ada ecommerce di miliki oleh negara..namun potongan nya jngn terlalu besar seperti milik swasta.slama ini ecommerce di indonesia di kuasai swasta dan investor asing..sudah saat nya ecommerce dimiliki oleh negara untuk mendukung umkm. selama ini sya ikut jualan di platform swasta sperti si ijo si biru si oren dan bahkan tok tok.nmun potongan nya sekarang terlalu besar.penjualan 25 rb.yg kita terima paling 17k setelah dipotong.17k itu kita cuma untung 3k aja..ayo bumN bikin segera.	['pengen', 'rasa', 'nya', 'ada', 'ecommerce', 'di', 'miliki', 'oleh', 'negara', 'namun', 'potongan', 'nya', 'jngn', 'terlalu', 'besar', 'seperti', 'milik', 'ini', 'ecommerce', 'di', 'indonesia', 'di', 'kuasai', 'swasta', 'dan', 'investor', 'asing', 'sudah', 'saat', 'nya', 'ecommerce', 'dimiliki', 'oleh', 'negara', 'untuk', 'mendukung', 'umkm', 'selama', 'ini', 'sya', 'ikut', 'jualan', 'di', 'platform', 'swasta', 'sperti', 'si', 'ijo', 'si', 'biru', 'si', 'oren', 'dan', 'bahkan', 'tok', 'potongan', 'nya', 'sekarang', 'terlalu', 'kita',

No	Sentimen sebelum <i>tokenizing</i>	Sentimen sesudah <i>tokenizing</i>
2	chat penjual tdk ada tanggapan krn offline sepanjang hr...akibatnya pesan loading nya g sampe sampe. jelek respon chatnya. sdh dibayar barang blm dikirim krn respon penjual dr chatnya jelek.sdh dibayar tgl 9 sampe dengan tanggal 14 pun tidak dikirim atau direfund uang nya. ini aplikasi penipuan. hati hati jangan transaksi dengan aplikasi ini	['terima', 'paling', 'setelah', 'itu', 'kita', 'cuma', 'untung', 'aja', 'ayo', 'bumn', 'bikin', 'segera'] ['chat', 'penjual', 'tdk', 'ada', 'tanggapan', 'krn', 'offline', 'sepanjang', 'hr', 'akibatnya', 'pesan', 'loading', 'nya', 'g', 'sampe', 'sampe', 'jelek', 'respon', 'chatnya', 'sdh', 'dibayar', 'barang', 'blm', 'dikirim', 'krn', 'respon', 'penjual', 'dr', 'chatnya', 'dibayar', 'tgl', 'sampe', 'dengan', 'tanggal', 'pun', 'tidak', 'dikirim', 'atau', 'direfund', 'uang', 'nya', 'ini', 'aplikasi', 'penipuan', 'hati', 'hati', 'jangan', 'transaksi', 'dengan', 'aplikasi', 'ini']
3	bagus digunakan	['bagus', 'digunakan']
4	server error terus, udah bayar ga kedetect!! pesanannya yg udh saya bayar malah tidak ada di history	['server', 'error', 'terus', 'udah', 'bayar', 'ga', 'kedetect', 'pesanannya', 'yg', 'udh', 'saya', 'bayar', 'malah', 'tidak', 'ada', 'di', 'history']
5	server gangguan terus, berat berat, padahal diwajibkan beli di sini	['server', 'gangguan', 'terus', 'berat', 'berat', 'padahal', 'di', 'wajibkan', 'beli', 'di', 'sini']
..
108	bermanfaat	['bermanfaat']
109	hidup umkm	['hidup', 'umkm']

3. *Stemming*

selanjutnya tahapan stemming yang bertujuan untuk membentuk teks menjadi kata dasar atau dikenal dengan istilah term. Proses ini dilakukan dengan menghapus imbuhan berupa awalan dan akhiran. Selanjutnya term tersebut akan didaftar dan diberikan bobot. Adapun output tahapan stemming dilihat pada Tabel 4.

Tabel 4. Proses *Stemming*

No	Sentimen sebelum <i>stemming</i>	Sentimen sesudah <i>stemming</i>
1	pengen rasa nya ada ecommerce di miliki oleh negara..namun potongan nya jngn terlalu besar seperti milik swasta.slama ini ecommerce di indonesia di kuasai swasta dan investor asing..sudah saat nya ecommerce dimiliki oleh negara untuk mendukung umkm. selama ini sya ikut jualan di platform swasta sperti si ijo si biru si oren dan bahkan tok tok.nmun potongan nya sekarang terlalu besar.penjualan 25 rb.yg kita terima paling 17k setelah dipotong.17k itu kita cuma untung 3k aja..ayo bumnn bikin segera.	['pengen', 'rasa', 'nya', 'ada', 'ecommerce', 'di', 'miliki', 'oleh', 'negara', 'namun', 'potongan', 'nya', 'jngn', 'terlalu', 'besar', 'seperti', 'milik', 'ini', 'ecommerce', 'di', 'indonesia', 'di', 'kuasai', 'swasta', 'dan', 'investor', 'asing', 'sudah', 'saat', 'nya', 'ecommerce', 'dimiliki', 'oleh', 'negara', 'untuk', 'mendukung', 'umkm', 'selama', 'ini', 'sya', 'ikut', 'jualan', 'di', 'platform', 'swasta', 'sperti', 'si', 'ijo', 'si', 'biru', 'si', 'oren', 'dan', 'bahkan', 'tok', 'potongan', 'nya', 'sekarang', 'terlalu', 'kita', 'terima', 'paling', 'setelah', 'itu', 'kita', 'cuma', 'untung', 'aja', 'ayo', 'bumnn', 'bikin', 'segera']
2	chat penjual tdk ada tanggapan krn offline sepanjang hr...akibatnya pesan loading nya g sampe sampe. jelek respon chatnya. sdh dibayar barang blm dikirim krn respon penjual dr chatnya jelek.sdh dibayar tgl 9 sampe dengan tanggal 14 pun tidak dikirim atau direfund uang nya. ini aplikasi penipuan. hati hati jangan transaksi dengan aplikasi ini	['chat', 'penjual', 'tdk', 'ada', 'tanggapan', 'krn', 'offline', 'sepanjang', 'hr', 'akibatnya', 'pesan', 'loading', 'nya', 'g', 'sampe', 'sampe', 'jelek', 'respon', 'chatnya', 'sdh', 'dibayar', 'barang', 'blm', 'dikirim', 'krn', 'respon', 'penjual', 'dr', 'chatnya', 'dibayar', 'tgl', 'sampe', 'dengan', 'tanggal', 'pun', 'tidak', 'dikirim', 'atau', 'direfund', 'uang', 'nya', 'ini', 'aplikasi', 'penipuan', 'hati', 'hati', 'jangan', 'transaksi', 'dengan', 'aplikasi', 'ini']
3	bagus digunakan	['bagus', 'digunakan']

No	Sentimen sebelum <i>stemming</i>	Sentimen sesudah <i>stemming</i>
4	server error terus, udah bayar ga kedetect!! pesanannya yg udh saya bayar malah tidak ada di history	['server', 'error', 'terus', 'udah', 'bayar', 'ga', 'kedetect', 'pesanannya', 'yg', 'udh', 'saya', 'bayar', 'malah', 'tidak', 'ada', 'di', 'history']
5	server gangguan terus, berat berat, padahal di wajibkan beli di sini	['server', 'gangguan', 'terus', 'berat', 'berat', 'padahal', 'di', 'wajibkan', 'beli', 'di', 'sini']
..
108	bermanfaat	['bermanfaat']
109	hidup umkm	['hidup', 'umkm']

4. TF-IDF

Pada tahap ini dikarenakan data yang dilabelkan menghasilkan imbalance data maka perlu dilakukan handling data, namun dikarenakan data berupa tipe data untaian atau string perlu dilakukan ekstraksi data menjadi tipe data numerik. TF-IDF menjadi metode pilihan untuk melakukan ekstraksi data menjadi tipe data numerik dikarenakan TF-IDF ini mengubah kata teks menjadi vektor numerik dengan menghitung frekuensi kemunculan kata dalam dokumen dan mengurangi bobot kata yang sering muncul dalam seluruh dokumen. Berikut hasil data yang sudah dilakukan TF-IDF dapat dilihat pada Tabel 5.

Tabel 5. Proses TF-IDF

	webnya	website	wilayah	wkwk	ya	yang	yg
Doc1	0	0	0	0	0	0	1
Doc2	0	0	0	0	0	0	0
Doc3	0	0	0	0	0	0	0
Doc4	0	0	0	0	0	0	1
Doc5	0	0	0	0	0	0	0
...
Doc108	0	1	0	0	0	0	0
Doc109	0	0	0	0	0	0	0
Doc110	0	0	0	0	0	0	0

3.3. Klasifikasi Menggunakan Algoritma KNN dan IKNN

Data yang sudah diolah pada tahapan sebelumnya, selanjutnya diimplementasikan pada usulan algoritma machine learning yaitu K-Nearest Neighbor atau KNN untuk melakukan klasifikasi user review pada Aplikasi PaDI UMKM dengan data testing yang digunakan sebanyak 22 data testing yaitu data ke 84, 2, 46, 13, 31, 97, 107, 26, 92, 6, 93, 85, 52, 62, 64, 9, 57, 87, 24, 49, 101, dan 29. Berikut hasil klasifikasi pengujian pada algoritma KNN dapat dilihat pada Tabel 6.

Tabel 6. Hasil Klasifikasi Menggunakan KNN dan IKNN

<i>Data Testing</i>	Klasifikasi KNN	Klasifikasi IKNN
ribet asli.. perhatikan kami yang jauh lokasinya, ongkir mahal.. masa uang habis untuk ongkir aja..	Positif	Negatif
bagus digunakan	Positif	Positif
saya mengajukan refund tapi gak segera dikirim uangnya, ini bagaimana ya aplikasi padi ini? kok begini lama banget dari mulai tgl 7 sampai sekarang tanggal 13 belum juga kembali uangnya	Negatif	Negatif
lemot, uninstal	Positif	Negatif
sangat membantu. serius! produk indonesia emang keren asli	Negatif	Positif
tidak bisa login, padahal id dan pasword sdh biasa dipakai saat login via web	Positif	Negatif
secara keseluruhan aplikasi bagus tapi mohon lebih di sederhanakan lagi agar masyarakat awam bisa menggunakan padi umkm seperti	Positif	Positif

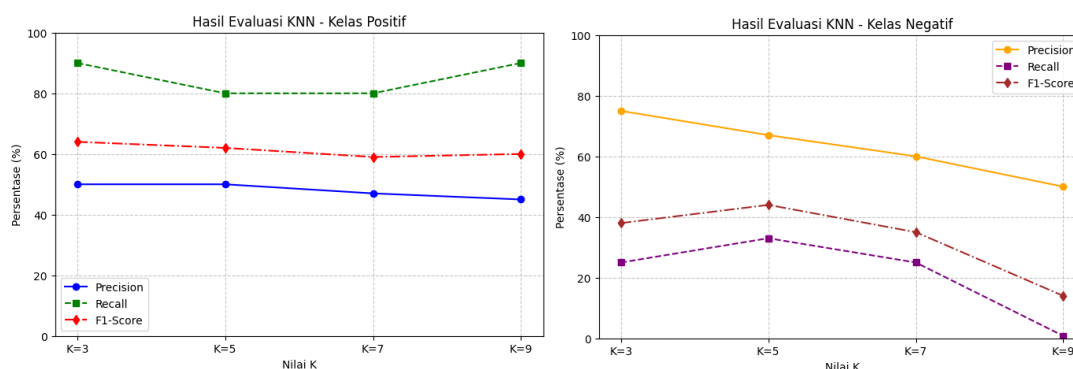
<i>Data Testing</i>	Klasifikasi KNN	Klasifikasi IKNN
toko kelontong dan toko grosir kecil dan pedagang kaki lima bisa masuk ke ekosistem digital padi umkm agar semua masyarakat indonesia bisa go digital untuk ekspor keluar negeri produk industri strategis indonesia, dan di aplikasi android tambahkan fitur daftar atau registrasi agar lebih mudah bergabung di padi umkm tanpa perlu berkunjung ke website padi umkm		
banyak promo menarik	Positif	Positif
cara jual barang gimana cara nya admin	Positif	Positif
susah belanja nya	Positif	Positif
kenapa gak di direct ke web browser aja kalau di suruh log in via web!?	Negatif	Negatif
ini gimana saya order pesanan kok ga dikirim2 ! aplikasi sampah. nipuuuuu	Negatif	Negatif
bantu toko saya untuk semuanya,siapa tau ada yg mau order pakaian/kaos di toko saya barang dijamin berkualitas	Positif	Positif
👍 👍	Positif	Positif
mantap	Positif	Positif
tidak bisa chat penjual, ktika di chat langsung keluar aplikasi	Positif	Negatif
kalau bukan karena diwajibkan perusahaan belanja disini ga bakalan sudi pake aplikasi ini!	Positif	Negatif
ini aplikasi gimana sih? login disuruh via web, giliran login di webnya mau verifikasi muter doang	Positif	Negatif
aplikasi gajelas	Positif	Negatif
sangat buruk. berbulan bulan gaji ada transaksi penjualan. buang buang memori. aplikasi buatan pemerintah emang gak ada yang beres	Negatif	Negatif
semoga bermanfaat dan dapat membantu kami para umkm	Negatif	Positif
aplikasi nggak jelas ajg, pesan dari bulan kemarin sampe sekarang nggak dikirim2, tau2 dah di cancel aja. hmmmmmm, plis jangan ada yang instal aplikasi ini.	Positif	Negatif

Berdasarkan hasil klasifikasi *user reviews* pada aplikasi PaDi UMKM dari 22 data testing didapatkan hasil sentimen menggunakan algoritma *machine learning* KNN menghasilkan klasifikasi ulasan negatif sebanyak 6 data dan ulasan postingan positif sebanyak 16 data. Sedangkan menggunakan algoritma IKNN ulasan positif sebanyak 10 data dan klasifikasi negatif pengguna sebanyak 12 data.

3.4. Evaluasi Menggunakan Algoritma KNN dan IKNN

Evaluasi dilakukan menggunakan confusion matrix yang diperoleh dari model yang dihasilkan. Berikut hasil nilai confusion matrix dari model algoritma KNN pada Tabel 7.

Tabel 7. Hasil Evaluasi Algoritma KNN				
Nilai K	Precision	Recall	F1-Score	Accuracy
K=3 (Positif)	50%	90%	64%	10%
K=3 (Negatif)	75%	25%	38%	12%
K=5 (Positif)	50%	80%	62%	10%
K=5 (Negatif)	67%	33%	44%	12%
K=7 (Positif)	47%	80%	59%	10%
K=7 (Negatif)	60%	25%	35%	12%
K=9 (Positif)	45%	90%	60%	10%
K=9 (Negatif)	50%	0,8%	14%	12%

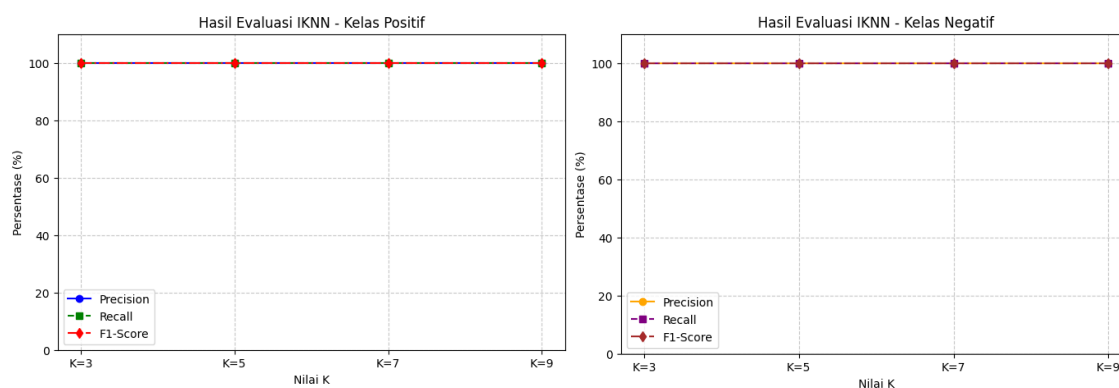


Gambar 2. Hasil Evaluasi KNN Kelas Positif dan Kelas Negatif

Berdasarkan hasil pengujian algoritma K-Nearest Neighbor (KNN) terhadap data ulasan pengguna aplikasi PaDi UMKM, diperoleh hasil yang menunjukkan bahwa performa model masih tergolong rendah. Pada kelas positif, nilai *Precision* berkisar antara 45% hingga 50%, sementara *Recall* cukup tinggi mencapai 90% pada beberapa nilai K, seperti pada K=3 dan K=9. Hal ini mengindikasikan bahwa model cukup baik dalam mengenali ulasan positif, namun belum mampu mempertahankan keseimbangan antara ketepatan dan kelengkapan prediksi, sebagaimana terlihat dari nilai *F1-Score* yang hanya mencapai sekitar 59%–64%. Sementara itu, pada kelas negatif, performa model jauh lebih rendah dengan *Recall* yang menurun drastis hingga 0,8% pada K=9 dan *F1-Score* terendah sebesar 14%. Secara keseluruhan, tingkat akurasi model yang hanya berkisar antara 10% hingga 12% menunjukkan bahwa algoritma KNN belum optimal dalam menganalisis sentimen terhadap aplikasi PaDi UMKM dan masih memerlukan peningkatan kinerja melalui pendekatan algoritma lain seperti Improved K-Nearest Neighbor (IKNN). Dengan demikian, hasil ini mengindikasikan perlunya peningkatan kinerja algoritma KNN, misalnya melalui penerapan metode Improved K-Nearest Neighbor (iKNN), untuk memperoleh hasil analisis sentimen yang lebih akurat dan representatif. Berikut hasil evaluasi menggunakan IKNN dapat dilihat pada Tabel 8.

Tabel 8. Hasil Evaluasi Algoritma IKNN

Nilai K	Precision	Recall	F1-Score	Accuracy
K=3 (Positif)	100%	100%	100%	100%
K=3 (Negatif)	100%	100%	100%	100%
K=5 (Positif)	100%	100%	100%	100%
K=5 (Negatif)	100%	100%	100%	100%
K=7 (Positif)	100%	100%	100%	100%
K=7 (Negatif)	100%	100%	100%	100%
K=9 (Positif)	100%	100%	100%	100%
K=9 (Negatif)	100%	100%	100%	100%



Gambar 3. Hasil Evaluasi IKNN Kelas Positif dan Kelas Negatif

Berdasarkan hasil pengujian menggunakan algoritma Improved K-Nearest Neighbor (IKNN) terhadap data ulasan pengguna aplikasi PaDi UMKM, diperoleh hasil yang menunjukkan performa model yang sangat optimal. Pada seluruh nilai K (3, 5, 7, dan 9), baik untuk kelas positif maupun negatif, nilai *Precision*, *Recall*, *F1-Score*, dan *Accuracy* semuanya mencapai 100%. Hal ini menunjukkan bahwa algoritma IKNN mampu mengklasifikasikan data sentimen dengan tingkat ketepatan dan konsistensi yang sempurna tanpa adanya kesalahan prediksi. Hasil ini juga memperlihatkan bahwa penambahan mekanisme perbaikan pada IKNN berhasil meningkatkan kinerja secara signifikan dibandingkan algoritma KNN standar, terutama dalam mendeteksi dan mengelompokkan ulasan positif maupun negatif dari pengguna aplikasi PaDi UMKM di Play Store secara akurat dan efisien.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan algoritma K-Nearest Neighbor (KNN) dan Improved K-Nearest Neighbor (IKNN) dalam analisis sentimen terhadap ulasan pengguna aplikasi PaDi UMKM menunjukkan perbedaan performa yang signifikan. Hasil pengujian menggunakan KNN dengan variasi nilai K (3, 5, 7, dan 9) memperlihatkan bahwa kinerja model masih tergolong rendah, dengan rata-rata akurasi hanya sekitar 10–12%. Nilai True Positive dan True Negative juga sangat kecil, masing-masing hanya sekitar 8–9% dan 2–3%, sementara False Negative mencapai sekitar 87–88%, yang menunjukkan bahwa model sering salah mengklasifikasikan ulasan positif menjadi negatif. Berdasarkan hasil confusion matrix, nilai K = 5 memberikan performa terbaik dibandingkan K = 3, K = 7, dan K = 9, dilihat dari peningkatan nilai Precision, Recall, F1-Score, dan Accuracy meskipun secara keseluruhan kinerjanya masih rendah. Kondisi ini menunjukkan bahwa algoritma KNN belum mampu mengenali pola sentimen dengan baik pada dataset yang digunakan.

Sebaliknya, penerapan algoritma IKNN menunjukkan peningkatan performa yang sangat signifikan, dengan nilai Precision, Recall, F1-Score, dan Accuracy mencapai 100% pada semua nilai K. Hasil ini membuktikan bahwa peningkatan metode KNN melalui pendekatan IKNN berhasil meningkatkan kemampuan model dalam mengelompokkan sentimen secara akurat dan konsisten. Dengan demikian, algoritma IKNN terbukti lebih efektif dan efisien dalam menganalisis sentimen pengguna terhadap aplikasi PaDi UMKM, serta dapat dijadikan acuan dalam penelitian selanjutnya untuk pengembangan sistem analisis sentimen berbasis kecerdasan buatan yang lebih adaptif dan presisi.

DAFTAR PUSTAKA

- [1] M. Febima, M. Asfi, C. Lukita, L. Magdalena, M. Hatta, R. Fahrudin, and S. Mita “Utilizing the Shopee Marketplace Platform to Optimize MSMEs Business Strategies in the International Digital Economy,” vol. 04, pp. 138–146, 2025.
- [2] L. Magdalena, M. Hatta, and M. Febima, “PENDAMPINGAN PENGEMBANGAN USAHA UMKM BARU DI KOTA CIREBON DENGAN EKOSISTEM MODEL BISNIS CANVAS,” vol. 4, no. 1, pp. 23–35, 2025.
- [3] E. Sera, H. Hazriani, M. Mirfan, and Y. Yuyun, “Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing,” *Pros. SISFOTEK*, pp. 237–243, 2023, [Online]. Available: <http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/406%0Ahttp://www.seminar.iaii.or.id/index.php/SISFOTEK/article/download/406/338>
- [4] N. Azriansyah, E. Indra, and N. Azriansyah, “Penerapan Natural Language Processing Untuk Analisis Sentimen Terhadap Aplikasi Streaming,” *J. Ilm. BETRIK (Besemah Teknol. Inf. dan Komputer)*, vol. 14, no. 2, pp. 273–282, 2023, [Online]. Available: <https://ejournal.pppmitpa.or.id/index.php/betrik/article/view/96>
- [5] M. Febima and L. Magdalena, “Predictive Analytics on Shopee for Optimizing Product Demand Prediction through K-Means Clustering and KNN Algorithm Fusion,” vol. 6, no. 2, pp. 751–765, 2024, doi: 10.51519/journalisi.v6i2.720.
- [6] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [7] M. Febima, L. Magdalena, M. Asfi, M. Hatta, and R. Fahrudin, “Implementasi Optimasi NLP dan KNN untuk User Review Aplikasi SAMPEAN Cirebon,” pp. 162–168.
- [8] S. G. Setyorini, Mustakim, “Application of The Nearest Neighbor Algorithm for Classification of Online Taxibike Sentiments In Indonesia In The Google Playstore Application Application of The Nearest Neighbor Algorithm for Classification of Online Taxibike Sentiments In Indonesia In”, doi: 10.1088/1742-6596/2049/1/012026.

-
- [9] M. Prasetya, M. Wulandari, and S. A. Nikmah, "Implementasi NLP (Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify," *Stain. (Seminar Nas. Teknol. Sains)*, vol. 3, no. 1, pp. 145–153, 2024.
 - [10] A. S. Ramadhan, L. Magdalena, and M. Febima, "Perbandingan Kinerja Algoritma K-Nearest Neighbor dan Naive Bayes untuk Klasifikasi Loyalitas Pelanggan (Studi Kasus : CV Cahaya Alam Indah)," vol. 4, no. 3, pp. 4831–4838, 2025.
 - [11] X. Li, J. Zhang, and F. Safara, "Improving the Accuracy of Diabetes Diagnosis Applications through a Hybrid Feature Selection Algorithm," *Neural Process. Lett.*, vol. 55, no. 1, pp. 153–169, 2023, doi: 10.1007/s11063-021-10491-0.
 - [12] C. Zai and A. R. Isnain, "Komparasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) pada Analisis Sentimen Capcut," pp. 272–285, 2024.